

Lorena Vicini

**Modelos de processo de Poisson não-
homogêneo na presença de um ou mais
pontos de mudança, aplicados a dados de
poluição do ar**

Campinas

2012

**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO
CIENTÍFICA**

LORENA VICINI


**Modelos de processo de Poisson não-homogêneo na
presença de um ou mais pontos de mudança,
aplicados a dados de poluição do ar**

**TESE DE DOUTORADO APRESENTADA AO INSTITUTO
DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO
CIENTÍFICA DA UNICAMP PARA OBTENÇÃO DO
TÍTULO DE DOUTORA EM ESTATÍSTICA**

**ORIENTADOR: LUIZ KOODI HOTTA
COORIENTADOR: JORGE ALBERTO ACHCAR**

**ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE
DEFENDIDA PELA ALUNA LORENA VICINI, E ORIENTADA PELO PROF.DR LUIZ KOODI HOTTA**


Assinatura do Orientador


Assinatura do Coorientador

CAMPINAS, 2012

FICHA CATALOGRÁFICA ELABORADA POR
ANA REGINA MACHADO - CRB8/5467
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA - UNICAMP

Vicini, Lorena
V663m Modelos de processo de Poisson não-homogêneo na presença de um ou mais pontos de mudança, aplicados a dados de poluição do ar / Lorena Vicini. – Campinas, SP : [s.n.], 2012.

Orientador: Luiz Koodi Hotta.
Coorientador: Jorge Alberto Achcar.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Poisson, Processo de. 2. Teoria bayesiana de decisão estatística. 3. Método de Monte Carlo. 4. Markov, Processos de. 5. Ar - Poluição - Métodos estatísticos. I. Hotta, Luiz Koodi, 1952-. II. Achcar, Jorge Alberto. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em inglês: Non-homogeneous Poisson process in the presence of one or more change-points, an application to air pollution data

Palavras-chave em inglês:

Poisson processes
Bayesian statistical decision theory
Monte Carlo method
Markov processes
Air - Pollution - Statistical methods

Área de concentração: Estatística

Titulação: Doutor em Estatística

Banca examinadora:

Luiz Koodi Hotta [Orientador]
Edwin Moises Marcos Ortega
Fernando Antonio Moala
Mário de Castro Andrade Filho
Ronaldo Dias

Data de defesa: 12-06-2012

Programa de Pós-Graduação: Estatística

Tese de Doutorado defendida em 12 de junho de 2012 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.



Prof(a). Dr(a). LUIZ KOODI HOTTA



Prof(a). Dr(a). EDWIN MOISÉS MARCOS ORTEGA



Prof(a). Dr(a). FERNANDO ANTONIO MOALA



Prof(a). Dr(a). MÁRIO DE CASTRO ANDRADE FILHO



Prof(a). Dr(a). RONALDO DIAS

Agradecimentos

Primeiramente agradeço a Deus por permitir-me concluir este trabalho, pela força e perseverança para atingir meus objetivos e por colocar em meu caminho pessoas que me ajudaram e me acompanharam nos momentos difíceis.

Agradeço ao meu orientador Luiz Koodi Hotta, por ter acreditado em mim, pelo apoio, disponibilidade, transmissão de conhecimento, que foram de fundamental importância durante todo o programa de doutorado e principalmente pelo exemplo como acadêmico de pesquisador e ser humano. Este exemplo eu irei levar por toda a vida.

Agradeço ao meu coorientador Jorge Alberto Achcar, pelo apoio, disponibilidade, pelas ideias que muito contribuíram neste trabalho. Agradeço também pela sua compreensão e disposição em transmitir seus conhecimentos diante das minhas dificuldades.

Aos professores do Programa de Pós-Graduação em Estatística do IMECC-UNICAMP, pelos seus ensinamentos que contribuíram de forma significativa em minha formação profissional.

Quero agradecer a toda minha família, em especial aos meus pais, Aldérico Vicini e Jandira Vicini, aos meus irmãos Mauricio e Everaldo, pela ajuda incondicional, pelas orações e pelo amor que sempre recebi. Gostaria de agradecer também *in memoriam* de Izaíra Vicini, pelas orações e apoio dado durante sua vida.

Quero agradecer em especial ao meu namorado Fidel, pela paciência, amor e por sempre ter estado presente nos momentos bons e difíceis, durante esse período.

Agradeço a chefia do departamento de estatística, professora Anaelena e professor Adriano, da Universidade Federal de Santa Maria, pelo apoio recebido na finalização

deste trabalho.

Agradeço também a todos os meus amigos que foram minha família durante todo este período.

Agradeço a CAPES pela bolsa emergencial durante o período de um ano. Agradeço principalmente a UNICAMP, pela bolsa de docência, PED A, a qual foi de fundamental importância para a conclusão dos meus estudos.

Gostaria também de agradecer ao Laboratório Epifisma.

Resumo

A poluição do ar é um problema que tem afetado várias regiões ao redor do mundo. Em grandes centros urbanos, como é esperado, a concentração de poluição do ar é maior. Devido ao efeito do vento, no entanto, este problema não se restringe a esses centros, e conseqüentemente a poluição do ar se espalha para outras regiões. Os dados de poluição do ar são modelados por processos de Poisson não-homogêneos (NHPP) em três artigos: dois usando métodos Bayesianos com *Markov Chain Monte Carlo* (MCMC) para dados de contagem, e um usando análise de dados funcionais. O primeiro artigo discute o problema da especificação das distribuições a priori, incluindo a discussão de sensibilidade e convergência das cadeias MCMC. O segundo artigo introduz um modelo incluindo pontos de mudança para NHPP com a função taxa modelada por uma distribuição gama generalizada, usando métodos Bayesianos. Modelos com e sem pontos de mudança foram considerados para fins de comparação. O terceiro artigo utiliza análise de dados funcionais para estimar a função taxa de um NHPP. Esta estimação é feita sob a suposição de que a função taxa é contínua, mas com um número finito de pontos de descontinuidade na sua primeira derivada, localizados exatamente nos pontos de mudança. A função taxa e seus pontos de mudança foram estimadas utilizando suavização *splines* e uma função de penalização baseada nos candidatos a pontos de mudança. Os métodos desenvolvidos neste trabalho foram testadas através de simulações e aplicados a dados de poluição de ozônio da Cidade do México, descrevendo a qualidade do ar neste centro urbano. Ele conta quantas vezes, em um determinado período, a poluição do ar excede um limiar especificado de qualidade do ar, com base em

níveis de concentração de ozônio. Observou-se que quanto mais complexos os modelos, incluindo os pontos de mudança, melhor foi o ajuste.

Abstract

Air pollution is a problem that is currently affecting several regions around the world. In major urban centers, as expected, the concentration of air pollution is higher. Due to wind effect, however, this problem does not remain constrained in such centers, and air pollution spreads to other regions. In the thesis the air pollution data is modeled by Non-Homogeneous Poisson Process (NHPP) in three papers: two using Bayesian methods with Markov Chain Monte Carlo (MCMC) for count data, and one using functional data analysis. Paper one discuss the problem of the prior specification, including discussion of the sensitivity and convergence of the MCMC chains. Paper two introduces a model including change point for NHPP with rate function modeled by a generalized gamma distribution, using Bayesian methods. Models with and without change points were considered for comparison purposes. Paper three uses functional data analysis to estimate the rate function of a NHPP. This estimation is done under the assumption that the rate function is continuous, but with a finite number of discontinuity points in its first derivative, exactly at the change-points. The rate function and its change-points were estimated using splines smoothing and a penalty function based on candidate change points. The methods developed in this work were tested using simulations and applied to ozone pollution data from Mexico City, describing the air quality in this urban center. It counts how many times, in a determined period, air pollution exceeds a specified threshold of air quality, based on ozone concentration levels. It was observed that the more complex the models, including change-points, the better the fitting.

Sumário

Introdução	1
1 Processos de Poisson não-homogêneos aplicados a dados de contagem: uma abordagem bayesiana considerando diferentes distribuições <i>a priori</i>	5
1.1 Modelos sem pontos de mudança	5
1.2 Inferência bayesiana	8
1.3 Aspectos computacionais	11
1.3.1 Resumo	14
2 Modelos de processo de Poisson não-homogêneo na presença de um ou mais pontos de mudança, aplicados a dados de poluição do ar	17
2.1 Modelo com pontos de mudança	17
2.2 Aspectos computacionais	20
2.2.1 Resumo	23
3 Estimando taxas do processo de Poisson não-homogêneo com múltiplos pontos de mudança usando <i>splines</i>	25

3.1	Introdução	25
3.1.1	Funções <i>Spline</i>	26
3.1.2	<i>B-spline</i>	27
3.2	Processo de Poisson não-homogêneo para dados funcionais	30
3.2.1	Suavização <i>Spline</i> e Operadores Diferenciais	31
3.2.2	Detectando pontos de mudança por equações diferen- ciais.	32
3.2.3	Resumo	35
4	Conclusão	37
4.1	Estudos futuros	40
	Referências Bibliográficas	41
5	Non-homogeneous Poisson processes applied to count data: a Bayesian approach considering different prior distributions	45
5.1	Introduction	46
5.2	Bayesian Inference	49
5.2.1	Prior Distributions	49
5.2.2	Posterior Distributions	50
5.3	Simulated study	52
5.4	An application with pollution data	58
5.5	Conclusions	63
6	Non-homogeneous Poisson process in the presence of one or more change-points: an application to air pollution data	67
6.1	Introduction	68

6.2	Description of the model	69
6.3	Bayesian inference	72
6.3.1	Models without change-points	73
6.3.2	Models with change-points	76
6.4	Application to simulated data	79
6.4.1	Simulated data without change-point	80
6.4.2	Simulated data with one change-point	83
6.4.3	Simulated data with two change-points	88
6.5	Application to ozone data	94
6.6	Final Remarks	98
7	Estimating rates of nonhomogeneous Poisson Process with multiples change- points using splines	107
7.1	Introduction	107
7.2	Nonhomogeneous Poisson Process Model for Functional Data	109
7.2.1	Spline functions	109
7.2.2	Spline Smoothing and Differential Operators	110
7.2.3	Detecting Change Points by Differential Equations	111
7.3	Simulated Example	113
7.4	Mexico City Ozone Pollution Data	119
7.5	Conclusions	129
.1	Estimating Differential Equation Parameters from Data	133
A	Análises adicionais dos capítulos: 5 e 6	137
A.1	Análise de sensibilidade de modelos com pontos de mudança	137
A.2	Figuras dados simulados, modelos sem pontos de mudança	138
A.3	Figuras dados simulados, modelos com um ponto de mudança	141
A.4	Figuras dados simulados, modelos com dois pontos de mudança	144

A.5	Figuras dados NE, modelos sem pontos de mudança	148
A.6	Figuras dados NE, modelos com um ponto de mudança	149

Introdução

A tese é composta de três artigos sobre aplicações de métodos estatísticos relacionados à modelagem de dados de poluição de ozônio. A análise desses dados é muito importante, pois pode auxiliar parte da população muito sensível, tais como os idosos, recém-nascidos e doentes, os quais podem sofrer uma deterioração em sua saúde quando expostos a níveis de ozônio acima de 0,11 partes por milhão (0,11ppm), por um período de uma hora ou superior à isso (vide, por exemplo, Achcar et al., 2009). Portanto, modelar é uma questão muito importante.

No primeiro artigo, “Non-homogeneous Poisson processes applied to count data: a Bayesian approach considering different prior distributions”, foi discutida a abordagem bayesiana para dados de contagem usando processos de Poisson não-homogêneos, considerando diferentes distribuições *a priori* para os parâmetros do modelo. Uma abordagem bayesiana usando métodos de simulação MCMC para esse modelo foi introduzida por Achcar et al. (1998) com uma aplicação a dados de confiabilidade de *software* e considerando distribuições *a priori* não-informativas para os parâmetros do modelo. Com as distribuições *a priori* não-informativas introduzidas por esses autores pode-se ter muitas dificuldades computacionais na utilização de métodos MCMC. Neste artigo foram consideradas diferentes distribuições *a priori* para os parâmetros do modelo proposto e estudando o efeito dessas distribuições sobre a convergência e precisão dos resultados obtidos. Para ilustrar a metodologia proposta são considerados dados simulados e um conjunto de dados referentes aos casos de poluição de uma região da Cidade do México. Com esta análise foi possível propormos distribuições *a priori* que melho-

rassem consideravelmente a convergência das cadeias MCMC do modelo.

No segundo artigo, “Modelos de processo de Poisson não-homogêneo na presença de um ou mais pontos de mudança, aplicados a dados de poluição do ar”, foi introduzido o modelo com pontos de mudança, para um processo de NHPP, com a função taxa modelada através da distribuição gama generalizada, sob o enfoque da análise bayesiana. Consideramos modelos com pontos de mudança. O trabalho utiliza a abordagem bayesiana com estimação dos parâmetros através de métodos MCMC. Aqui também discutimos a utilização de diferentes distribuições *a priori* para os parâmetros dos modelos, com estudos de convergência e sensibilidade da distribuição *a priori* dos mesmos. A utilização da função intensidade do NHPP, assumindo a forma da distribuição gama generalizada, utilizando inferência bayesiana para a estimação dos parâmetros com pontos de mudança mostrou-se uma excelente ferramenta para analisar dados de poluição do ar causados pelo gás ozônio. Foi possível observar com este trabalho que à medida que aplicamos modelos mais complexos, com a inclusão de mais parâmetros, as estimativas obtidas tendem a ser melhores, em relação a convergência das cadeias e o ajuste do modelo, do que em relação aos modelos mais simples, ou seja, sem pontos de mudança. Isso pode ser observado pela análise gráfica e também pelo critério de comparação de modelos DIC. Para ilustrar a metodologia proposta são considerados dados simulados e um conjunto de dados referentes aos casos de poluição de uma região da Cidade do México.

No terceiro artigo, “Estimating rates of nonhomogeneous Poisson Process with multiples change-points using *splines*”, utilizamos a análise de dados funcionais para estimar a função taxa de um processo de Poisson não-homogêneo. Neste trabalho foi assumido que a função taxa é uma função contínua, mas com um número finito de descontinuidades em sua primeira derivada, nos pontos de mudança. A função taxa e seus pontos de mudança foram estimados com suavização *splines* com uma penalização baseada nos candidatos ponto de mudança atual. Embora seja um problema em aberto, a construção de um algoritmo para encontrar o valor ótimo para o parâmetro suave mostramos que esta metodologia pode detectar o verdadeiro ponto

de mudança, independentemente da escolha do parâmetro de ajuste.

Aplicando ao conjunto de dados reais, verificamos que os dados de poluição da cidade do México são um processo não-homogêneo e as maiores mudanças, nas taxas, ocorrem apenas no primeiro ou segundo ano e qualquer outra mudança subsequente é suficientemente suave para ser capturada pela estrutura do *spline*.

Nos três trabalhos os métodos desenvolvidos foram testados por meio de simulação e utilizados para estudar os dados de poluição por gás ozônio da Cidade do México com base em níveis de concentração de ozônio que são ultrapassados num determinado período de tempo.

Todos os resultados para os três artigos, são obtidos utilizando o *software* R.

Este trabalho está organizado da seguinte forma. No capítulo 1 apresentamos modelos sem pontos de mudança para um processo de NHPP com a função taxa modelada através da distribuição gama generalizada sob o enfoque da análise bayesiana. São propostas diferentes distribuições *a priori* para os parâmetros deste modelo e é estudado o efeito dessas distribuições *a priori* na convergência e precisão dos resultados obtidos. No capítulo 2 introduzimos o modelo com pontos de mudança para um processo de NHPP com a função taxa também modelada através da distribuição gama generalizada, sob o enfoque da análise bayesiana. No capítulo 3 temos um modelo com pontos de mudança para um processo de NHPP com a função taxa modelada sob o enfoque da análise de dados funcionais. No capítulo 4 mostramos algumas observações e discussões dos resultados obtidos, bem como estudos futuros sobre este assunto abordado. No capítulo 5 apresentamos o artigo completo “Processos de Poisson não-homogêneos aplicados a dados de contagem: uma abordagem bayesiana considerando diferentes distribuições *a priori*”. No capítulo 6 apresentamos o artigo completo “Modelos de processo de Poisson não-homogêneo na presença de um ou mais pontos de mudança, aplicados a dados de poluição do ar”. No capítulo 7 apresentamos o artigo completo “Estimando taxas do processo de Poisson não-homogêneo com múltiplos pontos de mudança usando *splines*”. No apêndice A está disposto um anexo com parte das análises realizadas para desenvolver os capítulos 6 e 7.

Capítulo 1

Processos de Poisson não-homogêneos aplicados a dados de contagem: uma abordagem bayesiana considerando diferentes distribuições *a priori*

1.1 Modelos sem pontos de mudança

O processo de Poisson não-homogêneo tem sido aplicado em várias situações, como por exemplo na análise de dados de confiabilidade de *software*, dados de poluição do ar ou dados de contagens em estudos médicos. Na introdução do modelo utilizamos a mesma nomenclatura utilizada quando o modelo foi introduzido para análise de dados de confiabilidade de *software*. Denote por $M(t)$ o número acumulado de eventos no intervalo de tempo de $(0, t]$, para $t \geq 0$. $M(t)$ é modelado por um processo de Poisson não-homogêneo com função valor médio $m(t)$. A função valor médio $m(t)$ muda dependendo do modelo, ou seja, ela assume a forma da distribuição que está sendo utilizada. Pode-se também caracterizar a distribuição através da função intensidade $\lambda(t) = \frac{\partial m(t)}{\partial t}$.

Se $\lambda(t)$ é uma constante, de modo que $m(t)$ é linear então $M(t)$ é chamado de processo de Poisson homogêneo; caso contrário, ele é chamado de processo de Poisson não-homogêneo.

Muitas escolhas diferentes para a função $m(t)$ são consideradas na literatura, especialmente em confiabilidade de *software* (vide, por exemplo, Achcar et al., 1998). Goel e Okumoto (1979) assumiram que o número esperado de falhas de *software* até o tempo t , dado pela função valor médio $m(t)$, é não decrescente e limitada superiormente. Especificamente, eles consideraram a função média

$$m_1(t) = \theta (1 - e^{-\beta t}), \quad (1.1)$$

em que θ representa o número máximo de dias em que o padrão de qualidade do ar é violado por determinado poluente no intervalo $(0, T]$ e β é considerado a taxa em que os eventos ocorrem. Por (1.1), nós temos que $\lambda_1(t) = m_1'(t) = \theta\beta e^{-\beta t}$. Em Goel (1983) foi proposta uma generalização do modelo (1.1) dada pela função intensidade

$$\lambda_2(t) = \theta\beta\alpha t^{\alpha-1} e^{-\beta t^\alpha}, \quad (1.2)$$

a função média correspondente é

$$m_2(t) = \theta (1 - e^{-\beta t^\alpha}). \quad (1.3)$$

Observe que (1.1) e (1.3) podem ser escritos como um caso especial da forma geral em que a função valor médio é dada como

$$m(t) = \theta F(t), \quad (1.4)$$

em que $F(t)$ é uma função de distribuição acumulada.

Por outro lado, para qualquer função distribuição $F(t)$ temos um modelo válido.

Uma função distribuição muito utilizada, dada a sua grande flexibilidade de ajuste é a da distribuição gama generalizada. A distribuição gama generalizada possui, como casos particulares, distribuição Weibull, log-normal, gama, qui-quadrado, entre outras.

Utilizando a distribuição gama generalizada, temos que a função de média é

$$m_{GG}(t) = \theta I_k(\beta t^\alpha), \quad (1.5)$$

em que $I_k(s)$ é a integral da função gama incompleta dada por

$$I_k(s) = \frac{1}{\Gamma(k)} \int_0^s x^{k-1} e^{-x} dx. \quad (1.6)$$

De (1.6) obtemos a função intensidade dada por

$$\lambda_{GG}(t) = m'_{GG}(t) = \frac{1}{\Gamma(k)} \theta \beta^k \alpha t^{\alpha k - 1} e^{-\beta t^\alpha}. \quad (1.7)$$

Este modelo recebe o nome de gama generalizada. Quando k é inteiro nós podemos escrever $F(t)$ como

$$F(t) = 1 - e^{-\beta t^\alpha} \sum_{j=0}^{k-1} \frac{(\beta t^\alpha)^j}{j!}. \quad (1.8)$$

Os três modelos utilizados neste artigo são descritos da seguinte forma:

- a) Modelo I: todos os parâmetros são desconhecidos, de forma que $\lambda_{GG}(t)$ é dada por (1.7).
- b) Modelo II: com $k = 1$, $\lambda_{GG}(t)$ é dada por $\lambda_2(t)$ dada em (1.8). É o modelo de Goel (1983).
- c) Modelo III: com $k = 1$ e $\alpha = 1$, $\lambda_{GG}(t)$ se reduz a $\lambda_1(t)$, que é o modelo de Goel e Okumoto (1979).

Na análise destes modelos, podemos ter algumas dificuldades na obtenção das inferências bayesianas usando métodos de simulação MCMC. Para isso vamos estudar o efeito de diferentes distribuições *a priori* no desempenho dos algoritmos de simulação de amostras da distribuição *a posteriori* de interesse.

A idéia principal desse artigo é propor especificações de distribuições *a priori* diferentes para os parâmetros, pois quando temos o modelo com todos os parâmetros livres a serem estimados, foi verificado o problema de convergência. Devido a isso,

é realizada uma análise da sensibilidade na estimação dos parâmetros desse modelo em relação às distribuições *a priori*. A maior preocupação é com o parâmetro k , um parâmetro de forma da distribuição gama generalizada.

Neste artigo apresentaremos duas propostas de distribuições *a priori* para o parâmetro k do modelo gama generalizada. A primeira utiliza as distribuições *a priori* para os parâmetros utilizadas em Achcar et al. (1998). Dados os problemas com esta *a priori*, principalmente de convergência, é sugerida uma segunda distribuição *a priori* para o parâmetro k , uma distribuição exponencial truncada. Como veremos a seguir, esta distribuição mostra-se adequada. Outras distribuições foram testadas, mas não serão apresentadas. Optou-se por apresentar apenas as duas, pois a primeira é a sugerida na literatura e a segunda por apresentar bons resultados.

Para verificarmos a convergência das estimativas dos parâmetros, além de observarmos a convergência gráfica das cadeias de cada parâmetro também utilizamos a estatística de Gelman e Rubin (1992). Tal estatística se baseia na idéia de que as trajetórias de cadeias com diferentes valores iniciais deve ser a mesma após a convergência. A estatística de Gelman e Rubin será sempre maior que 1. A convergência pode ser avaliada pela proximidade do valor dessa estatística ao valor 1. Gelman (1996) sugere aceitar a convergência se a estatística for menor que 1.2. Caso a convergência seja alcançada, é possível fazer inferência sobre os parâmetros em questão.

1.2 Inferência bayesiana

Denote por $D_T = \{n; t_1, t_2, \dots, t_n; T\}$ o conjunto de dados, em que n é o número de eventos observados tais que $0 \leq t_1 < t_2 < \dots < t_n \leq T$, em que t_i são os dias em que o gás ozônio ultrapassa um determinado limiar no período de tempo $(0, T]$.

A função de verossimilhança para o vetor $\Theta = (\theta, \beta, \alpha, k)$ considerando o tempo

truncado do modelo (vide, por exemplo, Cox e Lewis, 1966) é dada por

$$L(\Theta | D_T) = \left(\prod_{i=1}^n \lambda(t_i) \right) \exp(-m(T)). \quad (1.9)$$

Em alguns casos é necessário inserir de uma variável latente, pois esta poderá servir de auxílio computacional. Considera-se a introdução da variável latente N' que tem distribuição Poisson com parâmetro $\theta[1 - F(T | \beta)]$ Achcar et al. (1998).

Considerando o modelo da distribuição gama generalizada dado em (1.5), a função de verossimilhança para o vetor de parâmetros $\Theta = (\theta, \alpha, \beta, k)$ é expressa como

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \theta \frac{\beta^k}{\Gamma(k)} \alpha t_i^{\alpha k - 1} e^{-\beta t_i^\alpha} \right\} \exp \left\{ - \int_0^T \theta \frac{\beta^k}{\Gamma(k)} \alpha u^{\alpha k - 1} e^{-\beta u^\alpha} du \right\},$$

tal que

$$L(\Theta | D_T) = \frac{\theta^n \alpha^n \beta^{kn}}{\{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} \exp \left\{ -\beta \sum_{i=1}^n t_i^\alpha - \theta I_k(\beta T^\alpha) \right\}. \quad (1.10)$$

Distribuições *a priori*

O primeiro conjunto de distribuições *a priori* para o modelo I são as distribuições sugeridas por Achcar et al. (1998), dadas por

- (i) $N' \sim \text{Poisson} \{ \theta [1 - I_k(\beta T^\alpha)] \}$;
- (ii) $\theta \sim \text{Gama}(a, b)$; a, b conhecido;
- (iii) $\beta \sim \text{Gama}(c, d)$; c, d conhecidos;
- (iv) $\alpha \sim \pi_1(\alpha)$ em que $\pi_1(\alpha) \propto \frac{1}{\alpha}$ para $\alpha (\alpha > 0)$;
- (v) $k \sim \pi_2(k)$ em que $\pi_2(k) \propto \frac{1}{k}$ para $k (k > 0)$.

No segundo conjunto, também para o modelo I, a modificação está na proposta das especificações *a priori* para os parâmetros α e k , que passam a ser

- (iv) $\alpha \sim \text{Gama}(e, f); e, f$ conhecidos ;
- (v) $k \sim \text{exponencial truncada}(a_n, g); a_n, g$ conhecidos .

$P(\lambda)$ denota a distribuição Poisson com parâmetro λ , e a, b, c, d, e, f e a_n são os hiperparâmetros conhecidos das distribuições gama, em que $\text{Gama}(a, b)$ denota uma distribuição gama com valor esperado $\frac{a}{b}$ e variância $\frac{a}{b^2}$.

Assumimos independência das distribuições *a priori* dos parâmetros. Os valores dos hiperparâmetros são dados nas aplicações. Como temos problemas de convergência nas cadeias do parâmetro k , nas aplicações são estudados os efeitos de outras prioris para este parâmetro.

Nos modelos restritos II e III são utilizadas as mesmas distribuições *a priori* para os parâmetros livres.

Distribuição *a posteriori*

A inferência será realizada com base nas informações fornecidas pela distribuição *a posteriori* dos parâmetros. Como a distribuição *a posteriori*, a distribuição *a priori* e a função de verossimilhança estão relacionadas, sendo que a função de verossimilhança para os processos de Poisson tanto para o caso homogêneo quanto para o caso não-homogêneo, pode ser dada como

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \lambda(t_i) \exp \left(- \int_{t_{i-1}}^{t_i} \lambda(x) dx \right) \right\} \exp \left\{ - \int_{t_n}^T \lambda(u) du \right\},$$

isto é,

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \lambda(t_i) \right\} \exp \left\{ - \int_0^T \lambda(u) du \right\}.$$

Logo, temos

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \lambda(t_i) \right\} \exp \{-m(T)\}. \quad (1.11)$$

Inicialmente vamos considerar o caso em que *a priori* de k é a sugerida por Achcar et al. (1998). A distribuição *a posteriori* é dada por

$$\begin{aligned} P(N', \alpha, \beta, \theta | D_T) &\propto \frac{\theta^{N'+n+a-1} \alpha^n \beta^{kn+c-1}}{N'! \{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k-1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} \\ &\times e^{-(b+1)\theta - (d + \sum_{i=1}^n t_i^\alpha)\beta} \pi_1(\alpha) \pi_2(k). \end{aligned} \quad (1.12)$$

Como a distribuição *a posteriori* em (1.12) não tem forma fechada, então recorre-se aos métodos MCMC para gerar amostras. O algoritmo utilizado para obter amostras da distribuição *a posteriori* em (1.12) é dado da seguinte forma. Para o subconjunto de parâmetros cuja densidade condicional completa é conhecida, amostra-se diretamente dela, utilizando o amostrador de Gibbs com passos de *Metropolis* e para o subconjunto de parâmetros nos quais a densidade condicional não é conhecida, amostra-se das mesmas utilizando passos de *Metropolis-Hastings* ver, (Metropolis et al., 1953; Hastings, 1970). Uma exposição dos métodos MCMC é dada em Casella e Berger (2001).

1.3 Aspectos computacionais

Aqui detalhamos o procedimento de amostragem dos parâmetros da distribuição *a posteriori* dada em (1.12).

Amostrando β

Conforme a distribuição *a posteriori* em (1.12), a distribuição condicional completa de β é dada por

$$P_\beta(\beta | \cdot) \propto \beta^{kn+c-1} e^{-\beta[\sum_{i=1}^n x_i^\alpha + d]} \{1 - I_k(\beta T^\alpha)\}^{N'}.$$

A núcleo de transição $q(\beta^{(j)}, \cdot) = \text{Gama}(\cdot; \beta^{(j)} \times u_2, u_2)$ foi utilizada na imple-

mentação do algoritmo de *Metropolis-Hastings*. Além disso, u_2 é selecionado de forma que a taxa de aceitação fique em torno de 40%. O movimento de $\beta^{(j)}$ para β^{prop} é aceito com probabilidade

$$\delta(\beta^{(j)}, \beta^{prop}) = \min \left\{ 1, \frac{P_\beta(\beta^{prop})q(\beta^{prop}, \beta^{(j)})}{P_\beta(\beta^{(j)})q(\beta^{(j)}, \beta^{prop})} \right\}.$$

Amostrando α

Conforme a distribuição *a posteriori* em (1.12), a distribuição condicional completa de α é dada por

$$P_\alpha(\alpha | \cdot) \propto \alpha^n \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} e^{-\beta \sum_{i=1}^n t_i^\alpha} \{1 - I_k(\beta T^\alpha)\}^{N'} \pi_1(\alpha).$$

A núcleo de transição $q(\alpha^{(j)}, \cdot) = \text{Gama}(\cdot; \alpha^{(j)} \times u_3, u_3)$ foi utilizada na implementação do algoritmo de *Metropolis-Hastings*. Além disso, u_3 é selecionado de forma que a taxa de aceitação fique em torno de 40%. O movimento de $\alpha^{(j)}$ para α^{prop} é aceito com probabilidade

$$\delta(\alpha^{(j)}, \alpha^{prop}) = \min \left\{ 1, \frac{P_\alpha(\alpha^{prop})q(\alpha^{prop}, \alpha^{(j)})}{P_\alpha(\alpha^{(j)})q(\alpha^{(j)}, \alpha^{prop})} \right\}.$$

Amostrando k

Conforme a distribuição *a posteriori* em (1.12), a distribuição condicional completa de k é dada por

$$P_k(k | \cdot) \propto \frac{\beta^{kn}}{\{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} \pi_2(k).$$

A núcleo de transição $q(k^{(j)}, \cdot) = \text{Gama}(\cdot; k^{(j)} \times u_4, u_4)$ foi utilizada na implementação do algoritmo de *Metropolis-Hastings*. Além disso, u_4 é selecionado de forma que a taxa de aceitação fique em torno de 40%. O movimento de $k^{(j)}$ para k^{prop} é aceito com probabilidade

$$\delta(k^{(j)}, k^{(prop)}) = \min \left\{ 1, \frac{P_k(k^{prop})q(k^{prop}, k^{(j)})}{P_k(k^{(j)})q(k^{(j)}, k^{prop})} \right\}.$$

Amostrando θ

Conforme a distribuição *a posteriori* em (1.12), a distribuição condicional completa de θ é dada por

$$P_\theta(\theta | \cdot) \sim Gama[a + n + N', b + 1]. \quad (1.13)$$

Amostrando N'

Conforme a distribuição *a posteriori* em (1.12), a distribuição condicional completa de N' é dada por

$$P_{N'}(N' | \cdot) \sim Poisson[\theta(1 - I_k(\beta T^\alpha))]. \quad (1.14)$$

Quando adotamos a especificação *a priori* do segundo conjunto, dada na equação 1.11, propostas no modelo I, temos que a distribuição *a posteriori* é dada por

$$\begin{aligned} P(N', \alpha, \beta, \theta | D_T) &\propto \frac{\theta^{N'+n+a-1} \alpha^{n+e-1} \beta^{kn+c-1}}{N'! \{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k-1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} \\ &\times e^{-(b+1)\theta - (d + \sum_{i=1}^n t_i^\alpha)\beta - f\alpha - a_n k} I_{\{0,g\}}(k), \end{aligned} \quad (1.15)$$

e as únicas modificações nas condicionais completas necessárias para o procedimento de amostragem ocorrem em (iv) e (v), que são substituídas por:

Amostrando α

Conforme a distribuição *a posteriori* em (1.15), a distribuição condicional completa de α é dada por

$$P_\alpha(\alpha | \cdot) \propto \alpha^{n+e-1} \left\{ \prod_{i=1}^n t_i^{\alpha k-1} \right\} e^{-\beta \sum_{i=1}^n t_i^\alpha - f\alpha} \{1 - I_k(\beta T^\alpha)\}^{N'}.$$

A núcleo de transição $q(\alpha^{(j)}, \cdot) = Gama(\cdot; \alpha^{(j)} \times u_5, u_5)$ foi utilizada na implementação do algoritmo de *Metropolis-Hastings*. Além disso, u_5 é selecionado de forma

que a taxa de aceitação fique em torno de 40%. O movimento de $\alpha^{(j)}$ para α^{prop} é aceito com probabilidade

$$\delta(\alpha^{(j)}, \alpha^{(prop)}) = \min \left\{ 1, \frac{P_\alpha(\alpha^{prop})q(\alpha^{prop}, \alpha^{(j)})}{P_\alpha(\alpha^{(j)})q(\alpha^{(j)}, \alpha^{prop})} \right\}.$$

Amostrando k

Conforme a distribuição *a posteriori* em (1.15), a distribuição condicional completa para k é dada por

$$P_k(k | \cdot) \propto \frac{\beta^{kn}}{\{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} e^{-a_n k} I_{\{0, g\}}(k).$$

A núcleo de transição $q(k^{(j)}, \cdot) =$ exponencial truncada $(\cdot; k^{(j)} \times u_6, u_6)$ foi utilizada na implementação do algoritmo de *Metropolis-Hastings*. Além disso, u_6 é selecionado de forma que a taxa de aceitação fique em torno de 40%. O movimento de $k^{(j)}$ para k^{prop} é aceito com probabilidade

$$\delta(k^{(j)}, k^{(prop)}) = \min \left\{ 1, \frac{P_k(k^{prop})q(k^{prop}, k^{(j)})}{P_k(k^{(j)})q(k^{(j)}, k^{prop})} \right\}.$$

1.3.1 Resumo

Neste artigo foi proposto realizar uma análise de sensibilidade, propondo várias especificações de distribuições *a priori* para os parâmetros do modelo introduzido previamente por Achcar et al. (1998). Aqui realizamos um estudo do efeito das distribuições *a priori* na convergência e precisão dos resultados obtidos. Foi possível com isso propor uma distribuição *a priori* para o parâmetro k , sendo que quando este parâmetro é estimado temos problemas de convergência dos parâmetros. Após testarmos várias especificações de distribuições *a priori* para este parâmetro, foi possível propormos uma distribuição *a priori* que melhorasse consideravelmente a convergência das cadeias dos parâmetros do modelo. Isso pode ser observado nos intervalos de credibilidade, nos quais a amplitude do intervalo é menor utilizando uma distribuição *a priori* exponencial

truncada. Podemos também observar esse melhor resultado pela análise gráfica, na qual quando utilizamos uma distribuição *a priori* exponencial truncada para o parâmetro k , obtemos a convergência das cadeias dos parâmetros.

O artigo encontra-se no capítulo 5.

Capítulo 2

Modelos de processo de Poisson não-homogêneo na presença de um ou mais pontos de mudança, aplicados a dados de poluição do ar

2.1 Modelo com pontos de mudança

A poluição de ozônio muitas vezes sofre alterações durante o intervalo de tempo $(0, T]$, devido a algum tipo de intervenção ou mudança. Por exemplo, uma decisão política, que pode ter como possível consequência uma diminuição ou um aumento das emissões de gás ozônio e, portanto, um efeito similar pode ocorrer nas medições diárias.

Nesse caso podemos ter pontos de mudança, que denotamos aqui por τ . Para explicar esse fenômeno, utiliza-se aqui, modelos do Processos de Poisson Não-homogêneos (NHPP).

Aqui consideramos o modelo com a presença de pontos de mudança. A análise será realizada com os modelos apresentados no capítulo anterior, sendo acrescentado à estes os pontos de mudança.

Assumimos aqui todos os parâmetros da distribuição gama generalizada desconhecidos, com $\Theta_i = (\theta_i, \alpha_i, \beta_i, k_i)$ e τ_j , em que $i = 1, \dots, J + 1$ e $j = 1, \dots, J$,

a serem estimados, sendo que o número de pontos de mudança podem ser (τ_{i-1}, τ_i) , em que $\tau_0 = 0$. Aqui fazemos inicialmente a inclusão de um ponto de mudança no modelo, num segundo e terceiro momento os demais são incluídos, sendo estes inseridos de forma análoga ao primeiro ponto. Neste caso, nós temos que a função taxa do processo Poisson, dada em Achcar et al. (2009), é da forma

$$\lambda(t | \Theta) = \left\{ \begin{array}{ll} \lambda(t | \Theta_1), & \text{se } 0 \leq t < \tau_1 \\ \lambda(t | \Theta_j), & \text{se } \tau_{j-1} \leq t < \tau_j, j = 2, 3, \dots, J, \\ \lambda(t | \Theta_{J+1}), & \text{se } \tau_J \leq t \leq T, \end{array} \right\} \quad (2.1)$$

em que $\lambda(t | \Theta_j), j = 1, 2, \dots, J + 1$ são as funções intensidade relacionadas com função de verossimilhança $L(\Theta | D_T) = \{\prod_{i=1}^n \lambda(t_i)\} \exp\{-m(T)\}$ e $\Theta_j, j = 1, 2, \dots, J + 1$ são os parâmetros associados ao NHPP entre os pontos de mudança.

Para o caso da distribuição gama generalizada a função taxa é da forma

$$\lambda(t | \Theta) = \frac{1}{\Gamma(k_j)} \theta_j \beta_j^{k_j} \alpha_j t^{\alpha_j k_j - 1} e^{-\beta_j t^{\alpha_j}}.$$

A função média $m(t | \Theta_j), j = 1, 2, \dots, J + 1$, dada em Achcar et al. (2009), é da forma

$$m(t | \Theta) = \left\{ \begin{array}{ll} m(t | \Theta_1), & \text{se } 0 \leq t < \tau_1 \\ m(\tau_1 | \Theta_1) + m(t | \Theta_2) - m(\tau_1 | \Theta_2), & \text{se } \tau_1 \leq t < \tau_2, \\ m(t | \Theta_{j+1}) - m(\tau_j | \Theta_{j+1}) \\ + \sum_{i=2}^j [m(\tau_i | \Theta_i) - m(\tau_{i-1} | \Theta_i)] + m(\tau_1 | \Theta_1), & \text{se } \tau_j \leq t < \tau_{j+1}, \\ j = 2, 3, \dots, J, & \end{array} \right. \quad (2.2)$$

sendo que toma-se $\tau_{J+1} = T$. Para o caso da distribuição gama generalizada a função média é da forma

$$m(t | \Theta) = \theta_j I_{k_j}(\beta_j t^{\alpha_j}),$$

em que

$$I_{k_j}(s) = \frac{1}{\Gamma(k_j)} \int_0^s x^{k_j-1} e^{-x} dx.$$

Seja $\mathbf{w} = (\Theta_1, \Theta_2, \dots, \Theta_{J+1}; \tau_1, \tau_2, \dots, \tau_j)$ onde $\Theta_i = (\theta_i, \beta_i, \alpha_i, k_i)$ em que $i = 1, \dots, J+1$ e $j = 1, \dots, J$. Seja N_T o número de eventos até T e N_{τ_j} o número de eventos até τ_j . Tendo assumido um modelo de Poisson não-homogêneo, temos que a função de verossimilhança é dada por

$$\begin{aligned} L(w | D_T) &\propto \prod_{i=1}^{N_{\tau_1}} \lambda(t_i | w_1) e^{-m(\tau_1 | w_1)} \left[\prod_{j=2}^J \left(\prod_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \lambda(t_i | w_j) e^{-[m(\tau_j | w_j) - m(\tau_{j-1} | w_j)]} \right) \right] \\ &\times \prod_{i=N_{\tau_J}+1}^{N_T} \lambda(t_i | w_{J+1}) e^{-[m(T | w_{J+1}) - m(\tau_J | w_{J+1})]}. \end{aligned} \quad (2.3)$$

Para o caso da distribuição gama generalizada temos que a função de verossimilhança é da forma

$$\begin{aligned} L(w | D_T) &\propto \prod_{i=1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)} \theta_i \beta_i^{k_i} \alpha_i d_i^{\alpha_i k_i - 1} e^{-\beta_i d_i^{\alpha_i}} e^{-\theta_i I_{k_i}(\beta_i \tau_j^{\alpha_i})} \prod_{i=N_{\tau_j}+1}^{N_T} \frac{1}{\Gamma(k_i)} \theta_i \beta_i^{k_i} \alpha_i d_i^{\alpha_i k_i - 1} \\ &\times e^{-\beta_i d_i^{\alpha_i}} e^{-[\theta_i I_{k_i}(\beta_i T^{\alpha_i}) - \theta_i I_{k_i}(\beta_i \tau_j^{\alpha_i})]}. \end{aligned}$$

Neste artigo inserimos uma variável latente, a qual será utilizada para auxílio computacional. Seja N'_i uma variável latente que tem distribuição de Poisson com parâmetro $[\theta_i(1 - I_{k_i}(\beta_i t_i^{\alpha_i}))]$ (vide, por exemplo, Achcar et al., 1998).

Distribuições *a priori* para o modelo com pontos de mudança

As distribuições *a priori* atribuídas para os parâmetros são as seguintes:

- (i) $\tau_j \sim \text{Uniforme}(f_j, g_j), j = 1, 2;$
- (ii) $N'_i \sim \text{Poisson}[\theta_i(1 - I_{k_i}(\beta_i t_i^{\alpha_i}))], i = 1, 2, 3;$
- (iii) $\theta_i \sim \text{Gama}(a_i, b_i), i = 1, 2, 3, a_i, b_i$ conhecidos;
- (iv) $\beta_i \sim \text{Gama}(c_i, d_i), i = 1, 2, 3, c_i, d_i$ conhecidos;

(v) $\alpha \sim \text{Gama}(e_i, h_i)$, $i = 1, 2, 3$, e_i, h_i conhecidos;

(vi) $k \sim$ exponencial truncada (m_i, p) , $i = 1, 2, 3$ e m_i, p conhecidos.

Também assumimos independência *a priori* entre os parâmetros.

Distribuição *a posteriori*

Seja $\Phi = (\mathbf{w}, N'_i)$, denotados pelo vetor Φ . A inferência será realizada usando informações fornecidas pela distribuição *a posteriori* dos parâmetros. Assumindo independência entre os parâmetros, nós temos que, na presença de pontos de mudança a distribuição *a posteriori*, a distribuição *a priori* e a função de verossimilhança são relacionadas da seguinte maneira:

$$P(\Phi \mid D_T) \propto L(D_T \mid \Phi)P(\Phi), \quad (2.4)$$

em que $P(\Phi \mid D_T)$ é a distribuição *a posteriori* conjunta de Φ condicionada aos dados D_T ; $P(\Phi)$ corresponde às distribuições *a priori* dos parâmetros e $L(D_T \mid \Phi)$ é a função de verossimilhança.

Como a distribuição *a posteriori* em (2.4) não tem forma fechada, recorre-se aos métodos MCMC para gerar amostras dela.

2.2 Aspectos computacionais

Aqui detalhamos o procedimento de amostragem do parâmetro Φ da distribuição *a posteriori* dada em (2.4).

Amostrando β_i

De acordo com a distribuição *a posteriori* em (2.4), a distribuição condicional completa de β_i é dada por

$$P_{\beta}(\beta_i \mid \cdot) \propto \beta_i^{K_i N_{\tau_j} + c_i - 1} e^{-\beta_i (\sum_{i=1}^{N_{\tau_j}} t_i^{\alpha_i} + d_i)} [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N'_i}.$$

A núcleo de transição $q(\beta_i^{(j)}, \cdot) = \text{Gama}(\cdot; \beta_i^{(j)} \times u_2, u_2)$ foi utilizada na implementação do algoritmo de *Metropolis-Hastings*. Além disso, u_2 é selecionado de forma que a taxa de aceitação fique em torno de 40%. O movimento de $\beta_i^{(j)}$ para β_i^{prop} é aceito com probabilidade

$$\delta(\beta_i^{(j)}, \beta_i^{prop}) = \min \left\{ 1, \frac{P_\beta(\beta_i^{prop})q(\beta_i^{prop}, \beta_i^{(j)})}{P_\beta(\beta_i^{(j)})q(\beta_i^{(j)}, \beta_i^{prop})} \right\}.$$

Amostrando α_i

De acordo com a distribuição *a posteriori* em (2.4), a distribuição condicional completa de α_i é dada por

$$P_\alpha(\alpha_i | \cdot) \propto \alpha_i^{N_{\tau_j} + e_i - 1} \left(\prod_{i=1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1} \right) e^{-\beta_i \sum_{i=1}^{N_{\tau_j}} t_i^{\alpha_i} - h_i \alpha_i} [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N'_i}.$$

A núcleo de transição $q(\alpha_i^{(j)}, \cdot) = \text{Gama}(\cdot; \alpha_i^{(j)} \times u_3, u_3)$ foi utilizada na implementação do algoritmo de *Metropolis-Hastings*. Além disso, u_3 é selecionado de forma que a taxa de aceitação fique em torno de 40%. O movimento de $\alpha_i^{(j)}$ para α_i^{prop} é aceito com probabilidade

$$\delta(\alpha_i^{(j)}, \alpha_i^{prop}) = \min \left\{ 1, \frac{P_\alpha(\alpha_i^{prop})q(\alpha_i^{prop}, \alpha_i^{(j)})}{P_\alpha(\alpha_i^{(j)})q(\alpha_i^{(j)}, \alpha_i^{prop})} \right\}.$$

Amostrando τ_j

De acordo com a distribuição *a posteriori* em (2.4), a distribuição condicional completa de α_i é dada por

$$\begin{aligned} P_\tau(\tau_j | \cdot) &\propto \left(\prod_{i=1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)} \right) \theta_i^{N_{\tau_j}} \beta_i^{k_i N_{\tau_j}} \alpha_i^{N_{\tau_j}} \left(\prod_{i=1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1} \right) e^{-\beta_i \sum_{i=1}^{N_{\tau_j}} t_i^{\alpha_i}} \left(\prod_{i=N_{\tau_j}+1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)} \right) \\ &\times \theta_i^{N_{\tau_j} - N_{\tau_j - 1}} \beta_i^{k_i N_{\tau_j} - N_{\tau_j - 1}} \alpha_i^{N_{\tau_j} - N_{\tau_j - 1}} \left(\prod_{i=N_{\tau_j}+1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1} \right) e^{-\beta_i \sum_{i=N_{\tau_j}+1}^{N_{\tau_j}} t_i^{\alpha_i}} \\ &\times [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N'_i} [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i}) + I_{k_i}(\beta_i \tau_{j-1}^{\alpha_i})]^{N'_i}. \end{aligned}$$

A núcleo de transição $q(\tau_i^{(l)}, \cdot) = \text{Gama}(\cdot; \tau_i^{(l)} \times u_5, u_5)$ foi utilizada na implementação do algoritmo de *Metropolis-Hastings*. Além disso, u_5 é selecionado de forma que a taxa de aceitação fique em torno de 40%. O movimento de $\tau_i^{(l)}$ para τ_i^{prop} é aceito com probabilidade

$$\delta(\tau_i^{(l)}, \tau_i^{prop}) = \min \left\{ 1, \frac{P_\tau(\tau_i^{prop})q(\tau_i^{prop}, \tau_i^{(l)})}{P_\tau(\tau_i^{(l)})q(\tau_i^{(l)}, \tau_i^{prop})} \right\}.$$

Amostrando k_i

De acordo com a distribuição *a posteriori* em (2.4), a distribuição condicional completa de k_i é dada por

$$P_k(k_i | \cdot) \propto \left(\prod_{i=1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)} \right) \beta_i^{k_i N_{\tau_j}} \left(\prod_{i=1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1} \right) e^{-m_i k_i} I_{(0,n)} k_i [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N'_i}.$$

A núcleo de transição $q(k_i^{(j)}, \cdot) = \text{exponencial truncada}(\cdot; k_i^{(j)} \times u_4, u_4)$ foi utilizada na implementação do algoritmo de *Metropolis-Hastings*. Além disso, u_4 é selecionado de forma que a taxa de aceitação fique em torno de 40%. O movimento de $k_i^{(j)}$ para k_i^{prop} é aceito com probabilidade

$$\delta(k_i^{(j)}, k_i^{prop}) = \min \left\{ 1, \frac{P_k(k_i^{prop})q(k_i^{prop}, k_i^{(j)})}{P_k(k_i^{(j)})q(k_i^{(j)}, k_i^{prop})} \right\}.$$

Amostrando θ_i

De acordo com a distribuição *a posteriori* em (2.4), a distribuição condicional completa de θ_i é dada por

$$P_\theta(\theta_i | \cdot) \sim \text{Gama}(N_{\tau_j} + N'_i + a_i - 1, 1 + b_i). \quad (2.5)$$

Amostrando N'_i

De acordo com a distribuição *a posteriori* em (2.4), a distribuição condicional com-

pleta de N'_i é dada por

$$P_{N'}(N'_i | \cdot) \sim \text{Poisson}[\theta_i(1 - I_{k_i}(\beta_i \tau_j^{\alpha_i}))]. \quad (2.6)$$

A inserção dos pontos de mudança nos modelos II e III é realizada de forma análoga ao modelo I.

2.2.1 Resumo

Neste trabalho consideramos o problema de modelar os dados de poluição da Cidade do México. Como os dados de contagem que estamos analisando podem ter um comportamento crescente ou decrescente dependendo das medidas de intervenção, estamos utilizando a função intensidade do NHPP, a qual pode assumir este comportamento com o passar do tempo. Além disso, podemos ter a presença de pontos de mudança para a função intensidade.

A estimação dos parâmetros é feita utilizando uma formulação bayesiana baseada em métodos MCMC. Testes de sensibilidade foram realizados às distribuições *a priori* e mostraram que as distribuições *a posteriori* são robustas, ou seja, o resultado não é influenciado pelas distribuições *a priori*.

Fizemos também uma análise de sensibilidade utilizando para os parâmetros diferentes especificações de distribuições *a priori*, com estudos de convergência e sensibilidade dos mesmos. Para ilustrar a metodologia proposta foram considerados dados simulados e um conjunto de dados referentes aos casos de poluição de uma região da Cidade do México.

A análise foi realizada em duas etapas: Inicialmente trabalhamos com os modelos sem a presença de pontos de mudança, os quais são de fundamental importância, pois a informação obtida com estes modelos nos serviu de suporte para os modelos mais complexos utilizando pontos de mudança. Devido a isso encontramos modelos muito precisos para os dados de ozônio da Cidade do México. A utilização da informação obtida nas fases anteriores da análise bayesiana foi importante para se obter a con-

vergência das cadeias MCMC.

A aplicação do modelo gama generalizada, utilizando inferência bayesiana para NHPP com pontos de mudança, mostrou-se uma excelente ferramenta para analisar dados de poluição do ar causados pelo gás ozônio. Foi possível observar com este trabalho que à medida que aplicamos modelos mais complexos, com a inclusão de mais parâmetros, as estimativas obtidas tendem a ser melhores do que em relação aos modelos mais simples, ou seja, sem pontos de mudança. Como pode ser observado pelos gráficos e também pelo critério de comparação de modelos DIC.

Foi observado que os pontos de mudança ocorrem após um período em que algumas ações ambientais foram tomadas pelo governo a fim de reduzir os índices de poluição. Por exemplo, de 1990 a 1998 foi reduzido o uso de veículos altamente poluentes e por volta deste período temos nos dados um ponto de mudança (Achcar et al., 2009). É importante salientar que de 1999 a 2002 uma série de medidas foram tomadas pelas autoridades ambientais do México com o objetivo de reduzir os níveis de ozônio, tais como fabricantes que foram estimulados a produzir carros com tecnologia moderna e limpa, sendo esses produzidos de forma ampla a partir de 2001. Devido a isso, por volta de 2002 tivemos outro ponto de mudança.

O artigo encontra-se no capítulo 6.

Capítulo 3

Estimando taxas do processo de Poisson não-homogêneo com múltiplos pontos de mudança usando *splines*

3.1 Introdução

A dados funcionais podem ser definidos como dados de alta dimensão, ou ainda de dados que são formados por medidas repetidas de um mesmo objeto. Por exemplo os dados que são registrados densamente ao longo do tempo, como por máquinas, são geralmente chamados de dados funcionais, ou seja, uma curva (ou função) observada por objeto em estudo.

A análise de dados funcionais fornece ferramentas para previsão e para estimação dos momentos, curvatura, derivadas e outras características da distribuição subjacente que tem os dados funcionais. Métodos baseados em *splines* assumem somente que os dados são uma realização de um processo estocástico em um espaço funcional suavizado. A referência completa para *splines* está no livro de Boor (2001) sob um ponto de vista de análise numérica. Wahba (1990), Ferraty e Vieu (2006) e Ramsay e Silverman (2005) estão focados no uso estatístico dos métodos baseados em *splines*, tais como suavização por *splines*, regressão não paramétrica funcional e análise de componentes principais.

Quando uma mudança abrupta de regime acontece na função taxa $\lambda(t)$ no ponto c , c será chamado um ponto de mudança. A análise não paramétrica do processo de Poisson não-homogêneo da função taxa $\lambda(t)$ foi proposta primeiro por Leemis (1991) e uma metodologia baseada em *splines* foi mostrada por Ramsay e Silverman (2005). A estimação de pontos de mudança, considerando o cenário paramétrico pode ser revista no trabalho de Achcar et al. (2009) e referências contidas neste documento. Nossa proposta é que se possa entender pontos de mudança no modelo baseado em *splines* como uma descontinuidade na primeira derivada de $\lambda(t)$, assim unificando essas duas abordagens para o problema. Nosso trabalho visa estimar simultaneamente os pontos de mudança do processo de Poisson não-homogêneo com suavização por *splines* e a própria função taxa.

3.1.1 Funções *Spline*

Os polinômios devido a sua estrutura simples e boas propriedades de aproximação eles são constantemente utilizados para aproximar funções. Sendo assim, o intervalo $I = [a, b]$ o qual é dividido em subintervalos menores da forma $[x_0, x_1], \dots, [x_k, x_{k+1}]$ é então um polinômio de grau p_i (sendo este de grau menor que a função que será suavizada) é usado para aproximação em cada subintervalo.

Esse procedimento gera uma função de aproximação polinomial por partes $s(\cdot)$, sendo que, $s(t) = p_i(t)$ em cada intervalo $[x_i, x_{i+1}]$, em que $i = 0, \dots, k$. Os valores do polinômio $x_0, x_1, \dots, x_k, x_{k+1}$ são chamados de nós (do inglês *knots*), sendo que x_0 e x_{k+1} são os nós exteriores do polinômio e os demais x_1, \dots, x_k são os interiores.

Segundo Souza (2008) no caso geral, as partes de polinômio $p_i(t)$ são constituídas independentemente umas das outras, e, portanto, não formam uma função contínua $s(t)$ em $[a, b]$. Isso não pode ser aceito caso deseje, aproximar uma função suave. Sendo assim, é necessário que as partes do polinômio sejam unidas suavemente nos nós interiores x_1, \dots, x_k e também que sejam deriváveis um certo número de vezes. O resultado disso, é uma função polinomial por partes, suave, chamada de função *spline*.

Um *spline* de ordem m (ordem = grau+1) com k nós interiores em x_1, \dots, x_k é qualquer função da forma

$$s(t) = \sum_{i=0}^{m-1} \rho_i t^i + \sum_{i=1}^k \delta_i (t - x_i)_+^{m-1}, \quad (3.1)$$

em que os coeficientes $\rho_0, \dots, \rho_{m-1}, \delta_1, \dots, \delta_k$ são números reais e, dada uma função v , a “função de potência truncada” de grau r é definida como

$$v_+^r = \begin{cases} v^r, & \text{se } v \geq 0 \\ 0, & \text{se } v < 0. \end{cases} \quad (3.2)$$

Dessa forma, uma função *spline* é uma combinação linear de $m + k$ funções base. Levando em conta o conjunto de nós interiores x_1, \dots, x_k as funções base são dadas por $1, t, t^2, \dots, t^{m-1}, (t - x_1)_+^{m-1}, \dots, (t - x_k)_+^{m-1}$.

Uma função *spline* de ordem m com k nós interiores em x_1, \dots, x_k satisfaz as condições:

- $s(t)$ é um polinômio por partes, de grau $m - 1$ em qualquer subintervalo $[x_i, x_{i+1})$ para $i = 0, \dots, k$;
- $s(t), D_s(t), \dots, D^{m-2}s(t)$ são funções contínuas em I .

O conjunto de funções *spline* de ordem m e nós interiores em x_1, \dots, x_k é chamado de espaço *spline* e é denotado por $S_m(x_1, \dots, x_k)$. Segundo Schumaker (1981), este espaço é um espaço linear de dimensão $m + k$ e os chamados *B-Splines* formam uma base de espaços *spline*. Os *B-splines* têm suporte compacto, ou seja, são não-nulos em um intervalo pequeno e zero fora dele, que é uma importante propriedade computacional.

3.1.2 *B-spline*

Os *B-splines* de ordem m é constituído de m pedaços de polinômio, no qual cada um possui grau $m - 1$. Esses pedaços de polinômio são unidos em pelos nós internos. A escolha do número ideal de nós tem sido motivo de pesquisa, pois poucos nós resultam

em uma curva muito suave, já muitos nós produzem uma curva sem quase nenhuma suavidade.

Um *B-spline* de ordem m é constituído de m pedaços de polinômio que se juntam em $m - 1$ nós internos, sendo que nos pontos de união, as derivadas de ordem até $m - 2$ são contínuas. O *B-spline* é positivo no domínio abrangido pelos $m + 1$ nós, fora disso ele é zero. Exceto nas fronteiras, ele sobrepõem $2(m - 1)$ pedaços de polinômio de seus vizinhos. Dado um certo t , m *B-splines* são não nulos.

A Figura 3.1 mostra um *B-spline* de grau 1 isolado com três nós em "x" e um *B-splines* de grau 1 com nós posicionados em "x".

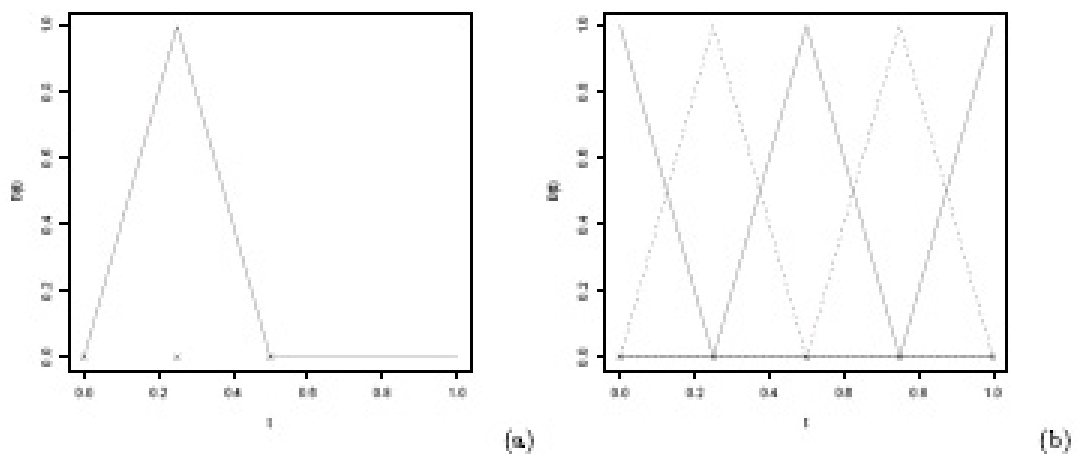


Figura 3.1: (a) *B-spline* de grau 1 isolado com três nós em "x". (b) *B-splines* de grau 1 com nós posicionados em "x", (vide, Souza 2008).

Segundo Souza (2008) um *B-spline* de grau 1 consiste de dois pedaços lineares, um pedaço de x_0 a x_1 , e outro de x_1 a x_2 . Os nós são x_0, x_1 e x_2 . À esquerda de x_0 e à direita de x_2 esse *B-spline* é zero ver Figura 3.1-(a). É possível construir um conjunto muito amplo de *B-splines*, para isso basta introduzir mais nós. Na Figura 3.1-(b) tem-se todos os *B-splines* possíveis de grau 1 no intervalo $[0, 1]$ com nós em $0; 0, 25; 0, 5; 0, 75; 1$.

Podemos definir um *B-spline* como uma diferença dividida da função potência truncada. Diferenças divididas surgem da interpolação polinomial de Newton, sendo que

$n + 1$ pares $t_i, f(t_i)_{i=0}^n$ um polinômio de grau n é aplicado da seguinte maneira:

$$p_n(t) = c_0 + c_1(t - t_0) + c_2(t - t_0)(t - t_1) + \cdots + c_n(t - t_0)(t - t_1) \cdots (t - t_{n-1}), \quad (3.3)$$

em que

$$\begin{aligned} p_n(t_0) &= c_0 = f(t_0) \\ p_n(t_1) &= c_0 + c_1(t_1 - t_0) = f(t_1) \\ p_n(t_2) &= c_0 + c_1(t_2 - t_0) + c_2(t_2 - t_0)(t_2 - t_1) = f(t_2), \end{aligned}$$

e assim por diante. O coeficiente c_0 é dado por $c_0 = f(t_0) = [t_0]f$, sendo $[t_0]f$ a diferença dividida de ordem zero da função $f(t)$ sobre o ponto t_0 . Sendo assim,

$$c_1 = \frac{f(t_1) - f(t_0)}{(t_1 - t_0)} = \frac{[t_1]f - [t_0]f}{t_1 - t_0} = [t_0, t_1]f,$$

sendo esta a diferença dividida de ordem 1 da função $f(t)$ sobre os pontos t_0 e t_1 . Dessa forma, podemos calcular todos os demais coeficientes $c_k, k = 0, \dots, n$ tal que

$$c_k = [t_0, \dots, t_k]f, \quad (3.4)$$

sendo essa a diferença dividida de ordem k da função $f(t)$ sobre os pontos t_0, \dots, t_k . Segundo Ruggiero e Lopes (1997) podemos mostrar que (3.4) é obtida de forma recursiva usando

$$[t_0, t_1, \dots, t_k]f = \frac{[t_1, \dots, t_k]f - [t_0, \dots, t_{k-1}]f}{t_k - t_0},$$

para $k = 1, \dots, n$, já que $[t_0]f = f(t_0)$. Logo uma função *spline* $s(t)$ pode ser escrita como uma combinação linear de funções base *B-spline* como

$$s(t) = \sum_{k=1}^K b_k B_k(t), \quad (3.5)$$

em que b_k são os coeficientes a serem determinados e $B_k(t)$ representa a j -ésima base B -spline avaliada no ponto t .

3.2 Processo de Poisson não-homogêneo para dados funcionais

Um dado funcional $y(t)$ é definido como uma realização de uma função aleatória $Y(t)$ tal que, para qualquer coleção de n pontos t_1, t_2, \dots, t_n , é caracterizada por

$$\mathbb{P}(Y(t_1) \leq y_1, \dots, Y(t_n) \leq y_n) = F_{t_1, t_2, \dots, t_n}(y_1, \dots, y_n).$$

Tal definição é devida a Yaglom (1987). Se assumirmos que

$$\mathbb{E}(Y(t)) = \alpha(t),$$

em que $\alpha(t)$ é uma função suavizada, então por uma estimativa do dado funcional nós pensamos na expansão de $\alpha(t)$ em uma combinação linear de K funções bases conhecidas $B_k(t)$, e respectivos coeficientes β_k , que são os parâmetros a serem estimados. Isto é,

$$\alpha(t) = \sum_{k=1}^K \beta_k B_k(t),$$

ou equivalentemente,

$$Y(t) = \sum_{k=1}^K \beta_k B_k(t) + \varepsilon(t),$$

sendo que nós assumimos $\varepsilon(t)$ ser um erro com média 0 e variância finita. Neste artigo utilizamos os B -splines como funções bases (Ramsay e Silverman, 2005). Dado um conjunto τ_1, \dots, τ_L de pontos em um intervalo fechado, chamados nós de ordem m , com B -spline $B_k(t)$, para $k = 1, \dots, K$, é um polinômio de ordem m em cada intervalo $[\tau_\ell, \tau_{\ell+1}]$, que é $m - 1$ vezes diferenciável nos outros nós τ_1 e τ_L , e pode ser obtido através

de uma relação recursiva (vide de Boor, 2001; Ramsay e Silverman, 2005). Eles formam a base para as funções *spline* no intervalo $[\tau_1, \tau_L]$.

A estimativa para a função taxa média $\lambda(t)$ através da expansão da função base para os processos de Poisson não-homogêneos foi mostrado em Ramsay e Silverman (2005, p.122). Tal estimação é um caso especial de estimação restrita, no sentido em que as estimativas finais devem ser funções não negativas. Para obter este resultado, as bases são simplesmente exponenciadas antes de serem introduzidas na função de verossimilhança do processo de Poisson. Em outras palavras, tomamos $\lambda(t)$ tal que

$$\lambda(t) = \exp \left\{ \sum_{k=1}^K \beta_k B_k(t) \right\} = \exp\{\boldsymbol{\beta}'\mathbf{B}(t)\}, \quad (3.6)$$

em que $B_k(t)$ são as bases *B-spline*, e β_k são os coeficientes a serem estimados (usamos a notação $\boldsymbol{\beta}'\mathbf{B}(t)$ como uma abreviatura para sua combinação linear daqui em diante). A função log-verossimilhança do processo de Poisson com taxa dada por (3.6) terá a forma

$$L(\boldsymbol{\beta} \mid t_1, t_2, \dots, t_n, T) = \sum_{i=1}^n \log(\lambda(t_i)) - \int_0^T \lambda(s) ds = \sum_{i=1}^n \boldsymbol{\beta}'\mathbf{B}(t_i) - \int_0^T \exp\{\boldsymbol{\beta}'\mathbf{B}(t)\} ds.$$

O problema é agora reduzido a encontrar $\boldsymbol{\beta}$ tal que L seja maximizada, que pode ser obtida numericamente.

3.2.1 Suavização *Spline* e Operadores Diferenciais

Os métodos de estimação *spline* possuem ferramentas interessantes tais como suavização e a possibilidade de ajustar modelos de equações diferenciais para os dados. Suavização robusta é discutida em Ramsay e Silverman (2005, p.81) e Wahba (1990), entre outros. Ela consiste em penalizar as estimativas com uma norma da segunda derivada grande, e o procedimento de estimação assim produz estimativas mais

suaves. Isto é, se definirmos

$$\text{PEN}_2(\lambda) = \int_{-\infty}^{\infty} \left[\frac{\partial^2}{\partial s^2} \lambda(s) \right]^2 ds, \quad (3.7)$$

então, podemos adicionar a penalização $\text{PEN}_2(\lambda)$, multiplicada por um parâmetro de suavização ξ para o critério de ajuste dos dados. Isto força as estimativas de β_k a produzir estimativas suaves de $\lambda(t)$. Embora existam muitas maneiras de encontrar o parâmetro de suavização ξ , não é fácil de minimizar com um critério GCV-like, como no caso regressão *splines* (detalhado em Wahba, 1990; Ramsay e Silverman, 2005). No entanto, escolhemos ξ de uma forma comparando a verossimilhança não penalizada contra a sua versão penalizada. Tal procedimento *naïve* segue as regras gerais sugeridas por Ramsay e Silverman (2005, p.95).

3.2.2 Detectando pontos de mudança por equações diferenciais.

O ajuste de equação diferencial é uma extensão natural do método de suavização, apresentado em Ramsay e Silverman (2005) e referências. Considere uma equação diferencial linear dada como

$$y^{(p)}(t) = -\gamma_0 y(t) - \gamma_1 y'(t) - \dots - \gamma_{(p-1)} y^{(p-1)}(t) + u(t), \quad (3.8)$$

que pode ser reescrita por

$$L_\gamma y(t) = \gamma_0 y(t) + \gamma_1 y'(t) + \dots + \gamma_{(p-1)} y^{(p-1)}(t) + y^{(p)}(t) - u(t) = 0. \quad (3.9)$$

Definimos $\text{PEN}_L(\lambda)$ por

$$\text{PEN}_L(\lambda) = \int_{-\infty}^{\infty} [L\lambda(s)]^2 ds,$$

em que L é um operador diferencial que anula a estrutura da equação diferencial, ou em outras palavras, $L\lambda^* = 0$ se, e somente se, λ^* é uma solução da equação diferencial (3.8). Esta penalização traz o estimador $\hat{\lambda}$ mais perto de ser uma das soluções λ^* . Hipóteses sobre a adequação dos métodos baseado em *splines* são discutidas em Lyche e Schumaker (1994). A existência de soluções para o caso geral é discutida em Ramsay et al. (2007), e nosso caso linear é discutida em Ramsay (2000). Teoria complementar para equações diferenciais baseadas em ajuste de dados funcionais pode ser encontrada no Apêndice do artigo que se encontra no capítulo 7 deste trabalho, em Ramsay et al. (2007), ou no capítulo 19 de Ramsay e Silverman (2005).

A equação diferencial cuja solução é dada por uma linha reta de inclinação γ_1 até o ponto de mudança c e depois tendo inclinação γ_2 , é dada por

$$\lambda'(t) = \gamma_1(1 - \Delta(t - c)) + \gamma_2\Delta(t - c),$$

em que $\Delta(t)$ é a função indicadora para o conjunto contínuo $\{t > 0\}$ e $\lambda'(t) = \frac{d}{dt}\lambda(t)$. Esta equação pode ser reescrita como

$$L_\gamma\lambda(t) = \lambda'(t) - \gamma_2\Delta(t - c) - \gamma_1(1 - \Delta(t - c)), \quad (3.10)$$

e a solução explícita para $L_\gamma\lambda(t) = 0$, sujeita à condição inicial $y(c) = \gamma_0$, é

$$\lambda(t) = \gamma_0 + \gamma_1(c - x)(1 - \Delta(x - c)) + \gamma_2(x - c)\Delta(x - c).$$

Nossa proposta é que podemos combinar essas técnicas para estimar a função taxa do processo de Poisson λ na presença de pontos de mudança. O ajuste da verossimilhança conjunta e o critério de fidelidade da equação diferencial é dado por

$$L_{L_\gamma}(\boldsymbol{\beta}, \xi \mid t_1, t_2, \dots, t_n, T) = \sum_{i=1}^n \boldsymbol{\beta}'\mathbf{B}(t_i) - \int_0^T \exp\{\boldsymbol{\beta}'\mathbf{B}(s)\} ds - \xi \int_0^T (L_\gamma\lambda(t))^2 dt. \quad (3.11)$$

Os estimadores de $\boldsymbol{\beta}$ e γ para um ξ fixo são obtidos por

$$\left(\hat{\boldsymbol{\beta}}_\xi, \hat{\gamma}_\xi\right) = \arg \max_{\boldsymbol{\beta}, \gamma} L_{L_\gamma}(\boldsymbol{\beta} \mid t_1, t_2, \dots, t_n, T),$$

e o estimador da taxa $\lambda(t)$ é

$$\hat{\lambda}(t) = \exp \left\{ \sum_{k=1}^K \hat{\beta}_k B_k(t) \right\}.$$

A primeira derivada da função taxa do processo de Poisson não-homogêneo, dado por *splines* exponenciadas, é dada por

$$\lambda'(t) = \left(\sum_{k=1}^K \boldsymbol{\beta}_k \mathbf{B}'_k(t) \right) \exp \left\{ \sum_{k=1}^K \boldsymbol{\beta}_k \mathbf{B}_k(t) \right\}, \quad (3.12)$$

e assim o operador (3.9) é explicitamente dado por

$$\|L_\gamma \lambda(t)\| = \int_{-\infty}^{\infty} \left(\left(\sum_{k=1}^K \boldsymbol{\beta}_k \mathbf{B}'_k(t) \right) \exp \left\{ \sum_{k=1}^K \boldsymbol{\beta}_k \mathbf{B}_k(t) \right\} - \gamma_2 \Delta(t-c) - \gamma_1 (1 - \Delta(t-c)) \right)^2 dt.$$

Tal penalização é a pedra angular do nosso trabalho: ela pode incorporar informação sobre o ponto de mudança c – se ele é conhecido – produzindo estimativas que se ajustam melhor aos dados, uma vez que a descontinuidade da primeira derivada poderia ser uma pequena violação da hipótese de suavidade do processo baseado em *splines* e, além disso, pode ser usada para encontrar um ou mais pontos de mudança desconhecidos pelo exame do critério de ajuste dos dados como o AIC (referimo-nos a Dias e Garcia, 2007, e suas referências para uma justificativa dos métodos baseados no AIC quando usamos *splines*, embora este artigo formule o problema de suavidade no contexto do número de base selecionadas K , em vez da norma PEN).

3.2.3 Resumo

Nossos métodos tentam expandir a teoria baseada em *splines* para processos de Poisson, relaxando a hipótese de suavidade em relação à verdadeira função $\lambda(t)$, incluindo uma penalização com o objetivo de capturar uma descontinuidade na derivada, enquanto a função se assume suave em outro lugar. A estimativa final ainda é uma função cúbica *spline*. Embora continue sendo um problema em aberto a construção de um algoritmo para encontrar o valor ótimo para o parâmetro suave, mostramos que esta metodologia pode detectar o verdadeiro ponto de mudança, independentemente da escolha parâmetro de ajuste.

Aplicando ao conjunto de dados reais, verificamos que os dados de poluição da cidade do México são um processo não-homogêneo e as maiores mudanças ocorrem apenas no primeiro ou segundo ano e qualquer outra mudança subsequente é suficientemente suave para ser capturada pela estrutura dos *splines*.

O artigo encontra-se no capítulo 7.

Capítulo 4

Conclusão

Nesta tese foram propostos alguns modelos probabilísticos para modelar os dados de poluição da Cidade do México.

Os modelos utilizados para desenvolver esta tese são dois sob o enfoque bayesiano e um utilizando análise de dados funcionais. Os resultados obtidos nos permitiram verificar o comportamento dos dados diários de poluição durante dezoito anos de observações (1990 a 2008), e com isso observar que as medidas tomadas para o controle da poluição na Cidade do México, estão sendo eficazes, ou seja, se os níveis de poluição diminuiram ao longo do período analisado. Foi possível observar também que à medida que aplicamos modelos mais complexos, com a inclusão de mais parâmetros, as estimativas obtidas tendem a ser melhores do que em relação aos modelos mais simples, isso pode ser observado pelo ajuste do modelo bem como pelo critério de comparação de modelos DIC.

No primeiro artigo foi proposto realizar uma análise de sensibilidade, propondo várias especificações de distribuições *a priori* para os parâmetros do modelo introduzido previamente por Achcar et al. (1998), o qual foi desenvolvido sob uma abordagem bayesiana. Foi realizado um estudo do efeito das distribuições *a priori* na convergência e precisão dos resultados obtidos. Conseguimos com isso, propor uma distribuição *a priori* para o parâmetro k , sendo que quando este parâmetro é estimado temos problemas de a

convergência das cadeias MCMC. Após testarmos várias especificações de distribuições *a priori* para este parâmetro, foi possível propormos uma distribuição *a priori* que melhorasse consideravelmente a convergência das cadeias MCMC do modelo. Isso pode ser observado nos intervalos de credibilidade, nos quais a amplitude do intervalo é menor utilizando uma distribuição *a priori* exponencial truncada. Foi possível também observarmos esse melhor resultado, pela análise gráfica, na qual quando utilizamos uma distribuição *a priori* exponencial truncada para o parâmetro k , obtemos a convergência das cadeias MCMC. Os resultados com dados simulados também foram constantes com essas conclusões.

No segundo artigo consideramos o problema de estimar o número de vezes que um padrão de qualidade do ar é ultrapassado num determinado período de tempo. Como os dados que estamos analisando podem ter um comportamento crescente ou decrescente dependendo das medidas de intervenção à poluição do ar, estamos utilizando a função intensidade do NHPP, a qual pode assumir esta forma com o passar do tempo. Além disso, podemos ter a presença de pontos de mudança para esta função. A estimação dos parâmetros foi feita utilizando uma formulação bayesiana baseada em métodos MCMC, sendo os resultados muito próximos das distribuições *a posteriori*. Testes de sensibilidade foram realizados nas distribuições *a priori* e mostraram que as distribuições *a posteriori* são robustas. Fizemos também uma análise de sensibilidade utilizando para os parâmetros diferentes especificações de distribuições *a priori*, com estudos de convergência e sensibilidade dos mesmos. Para ilustrar a metodologia proposta foram considerados dados simulados e um conjunto de dados referentes aos casos de poluição de uma região da Cidade do México. A análise foi realizada em duas etapas: Inicialmente trabalhamos com os modelos sem a presença de pontos de mudança, os quais são de fundamental importância, pois a informação obtida nesses modelos nos serviu de suporte para os modelos mais complexos, utilizando pontos de mudança. Devido a isso encontramos modelos muito precisos para os dados de ozônio da Cidade do México. A utilização da informação obtida nas fases anteriores da análise Bayesiana foi muito importante para obter a convergência das cadeias MCMC. A aplicação do modelo da

distribuição gama generalizada, utilizando inferência bayesiana para NHPP com pontos de mudança, mostrou-se uma excelente ferramenta, para analisar dados de poluição do ar causados pelo gás ozônio. Foi possível observar, com este trabalho que à medida que aplicamos modelos mais complexos, com a inclusão de mais parâmetros, as estimativas obtidas tendem a ser melhores do que em relação aos modelos mais simples, ou seja, sem pontos de mudança. Isso pode ser observado pela análise gráfica e também pelo critério de comparação de modelos DIC. Foi observado que os pontos de mudança ocorrem um período após que algumas ações ambientais tenham sido tomadas pelo governo a fim de reduzir os índices de poluição. Por exemplo, de 1990 a 1998, foi reduzido o uso de veículos altamente poluentes e por volta deste período temos no modelo um ponto de mudança (ver, Achcar et al., 2009). É importante salientar que de 1999 a 2002 uma série de medidas foram tomadas pelas autoridades ambientais do México, com o objetivo de reduzir os níveis de ozônio, tais como fabricantes que foram estimulados a produzir carros com tecnologia moderna e limpa, sendo esses produzidos de forma ampla em 2001. Devido a isso, por volta de 2002 tivemos outro ponto de mudança.

No terceiro artigo utilizamos métodos para tentar expandir a teoria baseada em *spline* para processos de Poisson, relaxando a hipótese de suavidade em relação à verdadeira função $\lambda(t)$, incluindo uma penalização que visa capturar uma descontinuidade na derivada, ainda suavizando a função em outro lugar. A estimativa final ainda é uma função cúbica *spline*. Embora continue sendo um problema em aberto a construção de um algoritmo para encontrar o valor ótimo para o parâmetro suave, mostramos que esta metodologia pode detectar o verdadeiro ponto de mudança, independentemente da escolha do parâmetro de ajuste. Na aplicação ao conjunto de dados reais, nós encontramos que, embora os dados de poluição da cidade do México sejam definitivamente um processo não-homogêneo, as maiores mudanças ocorrem apenas no primeiro ou segundo ano e qualquer outra mudança subsequente é suficientemente suave para ser capturada pela estrutura do *spline*.

4.1 Estudos futuros

É importante ressaltar que além dos modelos aqui apresentados, existem outros modelos que podem ser utilizados para analisar dados de poluição, como por exemplo, modelos de superposição de NHPP, os quais utilizam uma estrutura de independência entre dois NHPP, na qual as funções intensidade $\lambda_j(t)$ são exponenciais, podendo estas assumir também outras formas. Outro modelo que pode ser utilizado para analisar esses dados é a superposição de NHPP com análise de dados funcionais.

Outra forma de modelar esse processo é mediante a inserção de covariáveis nos modelos abordados neste trabalho.

A aplicação desses novos modelos é importante para que possamos compará-los aos modelos com apenas um NHPP, tendo por objetivo verificar quais desses modelos apresentam as melhores estimativas e ajustes desses dados.

Além de novos modelos é interessante aplicar outros critérios de seleção de modelos como o critério BIC (Schwarz, 1978) e o CPO (vide Gelfand et al., 1992).

Referências Bibliográficas

- [1] Achcar, J. A.; Dey, K.D. and Niverthi, M. *A Bayesian approach using nonhomogeneous Poisson process for software reliability models*, In: A.S.Basu; S.K.Basu; S.Mukhopadhyay, (Org.). *Frontiers in Reliability*. ed.1 River Edge, NJ: World Scientific, v. 4, p. 1–18, 1998.
- [2] Achcar, J. A.; Fernández-Bremauntz, A. A.; Rodrigues, E.R. E Tzintzun, G. Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model, *Environmetrics*, v. 19, p. 469–485, 2008.
- [3] Achcar, J. A.; Rodrigues, E.R.; and Tzintzun, G. – Using non-homogeneous Poisson models with multiple change-points to estimate the number of ozone exceedances in Mexico City, *Environmetrics*, v. 22, p. 1–12, 2011.
- [4] Boor, C. *A Practical Guide to Splines*, Revised Edition, New York: Springer, 2001.
- [5] Casella, G. and Berger, R. L. *Statistical Inference*, Second Edition, 2001.
- [6] Cox, D.R. and Lewis, P.A. *Statistical Analysis of Series of Events*, Methuen, 1966.
- [7] Dias, R. and Garcia, N. Consistent estimator for basis selection based on a proxy of the Kullback-Leibler distance, *Journal of Econometrics*, v. 141, p. 167–178, 2007.
- [8] Gelman, A. Inference and monitoring convergence. *Markov Chain Monte Carlo in Practice* (eds. W. R. Gilks, S. Richardson e D. J. Spiegelhalter), London: Chapman and Hall, p. 131–143, 1996.

- [9] Gelman, A. and Rubin, D. R. A single series from the gibbs sampler provides a false sense of security. *Bayesian Statistics 4* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith), Oxford: Oxford University Press, p. 625–631, 1992.
- [10] Gelfand , A. E.; Dey, D. K. e Chang, H. Model determination using predictive distribution with implementation sampling-based method (with discussion), *Technical Report*, p. 1–45, 1992.
- [11] Goel, A.L. and Okumoto, K. An Analysis Of Recurrent Software Errors In A Real-Time Control System, In Proceedings of ACM Conference, Washington, D.C., USA, p. 496–501, 1979.
- [12] Goel, A.L. A guidebook for software reliability assessment, *Technical Report*, 1983.
- [13] Hastings, W. K. –Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, v. 57, p. 97–109, 1970.
- [14] Kou, L, and Yang, T.Y. –Bayesian computation for nonhomogeneous Poisson process in software reliability, *To appear in Journal of the American Statistical Association*, 1995.
- [15] Lyche, T. and Schumaker, L. L. –L-Spline Wavelets, *Wavelets: Theory, Algorithms, and Applications*, p. 197–212, 1994.
- [16] Martins, L.C.; Latorre, M.R.D.O.; Cardoso, M.R.A.; Gonçalves, F.L.T.; Saldiva, P.H.N. e Braga, A.L.F. –Poluição Atmosférica e Atendimento por Pneumonia e Gripe em São Paulo, Brasil, *Revista de Saúde Pública*, v. 36, p. 88–94, 2002.
- [17] Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. –Equations of state calculations by fast computing machines, *Journal of Chemical Physics*, v. 21, p. 1087–1092, 1953.
- [18] Ramsay, J. O. –Differential Equation Models for Statistical Functions, *The Canadian Journal of Statistics*, v. 28, p. 225–240, 2000.

- [19] Ramsay, J. O. and Silverman, B. W. –*Functional Data Analysis*, Second Edition, New York: Springer-Verlag, 2005.
- [20] Ramsay, J. O.; Hooker, G.; Campbell, D. and Cao, J. –Parameter Estimation for Differential Equations: A Generalized Smoothing Approach, *Journal of the Royal Statistical Society*, p. 741–796, 2007.
- [21] Ruggiero, M. A. G.; Lopes, V. L. R. *Cálculo Numérico - Aspectos Teóricos e Computacionais*. 2. ed. [S.l.]: Pearson Makron Books, 1997.
- [23] Schwarz, G. Estimating the dimension of a model, *The Annals of Statistics*, v. 6, p. 461–464, 1978.
- [23] Schumaker, L. *Spline Functions: Basic Theory*. New York: Wiley, 1981.
- [24] Souza, C. P. E. de. *Testes de hipóteses para dados funcionais baseados em distâncias: um estudo usando splines*. Dissertação (Mestrado) - IMECC–UNICAMP, Campinas–SP, 2008.
- [25] Wahba, G. –*Spline Models for Observational Data*, p.169, 1990.
- [26] Yang, T.Y. –*Computational approaches to Bayesian inference for software reliability*, Ph.D. Thesis, University of Connecticut, Storrs, 1994.
- [27] Yaglom, A. M. –*Correlation Theory of Stationary and Related Random Functions* - New York: Springer-Verlag, 1987.

Capítulo 5

Non-homogeneous Poisson processes applied to count data: a Bayesian approach considering different prior distributions

Lorena Vicini, Luiz Koodi Hotta and Jorge Alberto Achcar

Abstract. This article aims at discussing the Bayesian approach for count data using non-homogeneous Poisson processes considering different prior distributions for model parameters. A Bayesian approach using Markov Chain Monte Carlo (MCMC) simulation methods for this model was firstly introduced by [1], taking into account software reliability data considering non-informative prior distributions for the parameters of the model. With the non-informative prior distributions presented by these authors, computational difficulties may occur when using MCMC methods. This article considers different prior distributions for the parameters of the proposed model and studies the effect of such priors on the convergence and accuracy of results found. In order to illustrate the proposed methodology two examples are considered: the first one considers simulated data, and the other one considers a set of data for pollution issues at a region located in Mexico City.

Keywords: Non-homogeneous Poisson processes, Bayesian analysis, Markov Chain Monte Carlo methods and simulation, prior distributions

5.1 Introduction

The non-homogeneous Poisson model has been applied to various situations, such as to software reliability data analysis, air pollution data or medical count data. When presenting the model, we will use the same nomenclature used when the model was presented for software reliability data analysis. Denote by $M(t)$ the cumulative number of events in the time interval $(0, t]$ for $t \geq 0$. $M(t)$ is modeled by a non-homogeneous Poisson process with mean value function $m(t)$. The mean value function $m(t)$ changes according to the model, ie, it assumes the shape of the distribution that is being used. One can also characterize the distribution by the intensity function $\lambda(t) = \frac{\partial m(t)}{\partial t}$.

If $\lambda(t)$ is a constant, so that $m(t)$ is linear then $M(t)$ is called a homogeneous Poisson process, otherwise it is called a non-homogeneous Poisson process.

Many different choices for the function $m(t)$ are considered in the literature, especially in software reliability (see, eg, Achcar et al., 1998). (Goel and Okumoto, 1979) considered that the expected number of software failures for time t , given by the mean value function $m(t)$ is nondecreasing and bounded above. Specifically, they have considered the mean function

$$m_1(t) = \theta (1 - e^{-\beta t}), \quad (5.1)$$

where θ represents the expected maximum number of days in which air quality standard is violated by a particular pollutant and β is considered the rate at which events occur. By (5.1), we have that $\lambda_1(t) = m'_1(t) = \theta\beta e^{-\beta t}$. In Goel (1983) a generalization of the model (5.1) was proposed, given by the following intensity function

$$\lambda_2(t) = \theta\beta\alpha t^{\alpha-1} e^{-\beta t^\alpha}, \quad (5.2)$$

The corresponding mean function is

$$m_2(t) = \theta (1 - e^{-\beta t^\alpha}). \quad (5.3)$$

Note that (5.1) and (5.3) can be written as special cases of the general form where the mean value function is given as

$$m(t) = \theta F(t), \quad (5.4)$$

where $F(t)$ is a distribution function. On the other hand, for any distribution function $F(t)$ we have a valid model.

A widely used distribution function, given its high flexibility of fit it is the Generalized Gamma distribution. In this case the mean function is

$$m_{GG}(t) = \theta I_k(\beta t^\alpha), \quad (5.5)$$

where $I_k(s)$ is the integral of the Gamma function given by

$$I_k(s) = \frac{1}{\Gamma(k)} \int_0^s x^{k-1} e^{-x} dx. \quad (5.6)$$

From (5.6), we obtain the intensity function given by

$$\lambda_{GG}(s) = m'_{GG}(t) = \frac{1}{\Gamma(k)} \theta \beta^k \alpha t^{\alpha k - 1} e^{-\beta t^\alpha}. \quad (5.7)$$

This model is called Generalized Gamma. When k is integer we can write $F(t)$ as

$$F(t) = 1 - e^{-\beta t^\alpha} \sum_{j=0}^{k-1} \frac{(\beta t^\alpha)^j}{j!}. \quad (5.8)$$

The three models used in this article are described as follows:

- a) Model I: Here all parameters are unknown, $\lambda_{GG}(t)$ is given by (5.7).
- b) Model II: with $k = 1$, $\lambda_{GG}(t)$ is given by $\lambda_2(t)$ given in (5.2) Goel (1983).

c) Model III: with $k = 1$ and $\alpha = 1$, $\lambda_{GG}(t)$ reduces to $\lambda_1(t)$ Goel and Okumoto (1979).

In the Bayesian analysis of these models, we may have some difficulties in obtaining the Bayesian inferences using MCMC simulation methods. In order to do so, let us study the effect of different prior distributions on the performance of sample simulation algorithms of the posterior distribution of interest.

The focus of this paper is to propose different prior distribution specifications for the parameters, because when we have the model with all free parameters to be estimated, was found the problem of convergence. Therefore, a sensitivity analysis in the estimating of this model's parameters is conducted to what regards prior distributions. The biggest concern is with the parameter k , a parameter of the gamma distribution.

In this article we will present two prior proposals to the parameter k of the Generalized Gamma model. The first one uses the prior distributions for the parameters used in Achcar et al. (1998). Given the problems with this prior, especially of convergence, a second prior distribution is suggested for the parameter k , a truncated exponential distribution. As we shall see, this distribution proved to be adequate. Other distributions were tested, but will not be displayed. We chose to present only the two distributions, because the first one is suggested in the literature and the second one has shown good results.

The convergence of convergence of the MCMC algorithm was analyzed by graphical method and by the Gelman-Rubin statistics. The Gelman-Rubin statistics (1992) relies on parallel chains to test whether they all converge to the same posterior distribution. Brooks and Gelman (1997) suggested the introduction of a correction factor. This statistics is evaluated using the coda package in R. This statistics is larger or equal to 1.0. The closest this statistics is to 1.0, more evidence we have that the chain is near convergence. The limit value of 1.2 is sometimes used as a guideline for "approximate convergence" (Gelman, 1996).

This article is outlined as follows. In section 5.2 we present the Bayesian inference for the models discussed here. Section 5.3 is an example of this process with simulated

data, and we will use the two prior distributions mentioned previously for the parameter k . An application to Mexico City's pollution data is performed in section 5.4. In section 5.5 some concluding remarks are made.

5.2 Bayesian Inference

Denote by $D_T = \{n; t_1, t_2, \dots, t_n; T\}$, the data set, where n is the number of events observed such that $0 \leq t_1 < t_2 < \dots < t_n \leq T$, where t_i are the times of the events observed during the period of time $(0, T]$.

The likelihood function for the vector $\Theta = (\theta, \beta, \alpha, k)$ considering the truncated time of the model (see, for example, Cox and Lewis, 1966) is given by

$$L(\Theta | D_T) = \left(\prod_{i=1}^n \lambda(t_i) \right) \exp(-m(T)). \quad (5.9)$$

In some cases it is necessary to enter a latent variable, as this may serve as a computational aid. We considered the introduction of the latent variable N' that has Poisson distribution with parameter $\theta[1 - F(T | \beta)]$ Achcar et al. (1998).

Considering the model of the generalized gamma distribution given in (5.5), the likelihood function for the parameters θ, α, β and k is expressed as

$$L(\theta, \alpha, \beta, k | D_T) = \left\{ \prod_{i=1}^n \theta \frac{\beta^k}{\Gamma(k)} \alpha t_i^{\alpha k - 1} e^{-\beta t_i^\alpha} \right\} \exp \left\{ - \int_0^T \theta \frac{\beta^k}{\Gamma(k)} \alpha u^{\alpha k - 1} e^{-\beta u^\alpha} du \right\},$$

such that

$$L(\theta, \alpha, \beta, k | D_T) = \frac{\theta^n \alpha^n \beta^{kn}}{\{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} \exp \left\{ -\beta \sum_{i=1}^n t_i^\alpha - \theta I_k(\beta T^\alpha) \right\}. \quad (5.10)$$

5.2.1 Prior Distributions

The first set of prior distributions for the general model I are the distributions suggested by Achcar et al. (1998):

- (i) $N' \sim \text{Poisson} \{\theta[1 - I_k(\beta T^\alpha)]\}$;
- (ii) $\theta \sim \text{Gamma}(a, b)$; a, b known ;
- (iii) $\beta \sim \text{Gamma}(c, d)$; c, d known ;
- (iv) $\alpha \sim \pi_1(\alpha)$ where $\pi_1(\alpha) \propto \frac{1}{\alpha}$ to $\alpha(\alpha > 0)$;
- (v) $k \sim \pi_2(k)$ where $\pi_2(k) \propto \frac{1}{k}$ to $k(k > 0)$.

In the second set, also for the general model I, the change is in the proposition of prior specifications for the parameters α and k , ie, given by

- (iv) $\alpha \sim \Gamma(e, f)$; e, f known ;
- (v) $k \sim \text{truncated exponential}(a_n, g)$; a_n, g known.

$P(\lambda)$ denotes the Poisson distribution with parameter λ , and a, b, c, d, e, f and a_n are the known hyperparameters of the gamma distribution where $\Gamma(a, b)$ denotes a gamma distribution with mean $\frac{a}{b}$ and variance $\frac{a}{b^2}$.

Let us assume prior independence for the parameters. The values of the hyperparameters are given in the applications. As we have convergence problems in the chains of the parameter k , in the applications the effects of other priors for this parameter are studied. These priors are given in section 5.3.

In restricted models II and III the same prior distributions for the free parameters are used.

5.2.2 Posterior Distributions

The inference will be conducted based on information supplied by the posterior distribution of the parameters. As a posterior distribution, the prior distribution, and likelihood function are related, and the likelihood function for the Poisson processes,

both for the homogeneous case, and for the non-homogeneous case, can be given as

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \lambda(t_i) \exp \left(- \int_{t_{i-1}}^{t_i} \lambda(x) dx \right) \right\} \exp \left\{ - \int_{t_n}^T \lambda(u) du \right\},$$

ie,

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \lambda(t_i) \right\} \exp \left\{ - \int_0^T \lambda(u) du \right\}.$$

Thus, we have

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \lambda(t_i) \right\} \exp \{-m(T)\}. \quad (5.11)$$

Initially we will consider the case where the prior of k suggested by Achcar et al. (1998) is assumed. The posterior distribution is given by

$$\begin{aligned} P(N', \alpha, \beta, \theta / D_T) &\propto \frac{\theta^{N'+n+a-1} \alpha^n \beta^{kn+c-1}}{N'! \{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k-1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} \\ &\times e^{-(b+1)\theta - (d + \sum_{i=1}^n t_i^\alpha)\beta} \pi_1(\alpha) \pi_2(k). \end{aligned} \quad (5.12)$$

Because the posterior distribution in (5.12) has no closed form, we resort to MCMC methods to generate samples. The algorithm used to obtain posterior distribution samples in (5.12) is given as follows. For the subset of parameters whose full conditional density is known, is sampled directly from it, using the Gibbs Sampling algorithm; and for the subset of parameters in which the conditional density is not known, the samples are taken using steps of the Metropolis-Hastings algorithm see, (Metropolis et al. 1953, Hastings 1970). Recommended reference for the review of MCMC methods are given in (Casella e Berger 2001).

The marginal posterior densities required for MCMC methods are given by

- (i) $N' | \theta, \alpha, \beta, k, D_T \sim P[\theta(1 - I_k(\beta T^\alpha))];$
- (ii) $\theta | N', \alpha, \beta, k, D_T \sim \Gamma[a + n + N', b + 1];$
- (iii) $\beta | N', \alpha, \theta, k, D_T \propto \beta^{kn+c-1} e^{-\beta[\sum_{i=1}^n x_i^\alpha + d]} \{1 - I_k(\beta T^\alpha)\}^{N'};$

$$(iv) \alpha | N', \theta, \beta, k, D_T \propto \alpha^n \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} e^{-\beta \sum_{i=1}^n t_i^\alpha} \{1 - I_k(\beta T^\alpha)\}^{N'} \pi_1(\alpha);$$

$$(v) k | \alpha, N', \theta, \beta, D_T \propto \frac{\beta^{kn}}{\{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} \pi_2(k).$$

The parameter θ have closed form distribution, so their posterior distributions are obtained through the method of Gibbs sampling. Since the posterior distributions of α, β and k have no closed form, so we resort to MCMC methods to approach them. The algorithm used to obtain posterior distribution samples of parameters α, β and k , whose full conditional density is not known it is sampled using Metropolis-Hastings algorithm steps.

When we adopt the prior specification of the second set, proposed in model I, we have that the posterior distribution is given by

$$P(N', \alpha, \beta, \theta / D_T) \propto \frac{\theta^{N'+n+a-1} \alpha^{n+e-1} \beta^{kn+c-1}}{N'! \{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} \\ \times e^{-(b+1)\theta - (d + \sum_{i=1}^n t_i^\alpha)\beta - f\alpha - a_n k} I_{\{0,g\}}(k), \quad (5.13)$$

and the only changes in the marginal posterior densities occur in (iv) and (v), which are replaced by

$$(iv) \alpha | N', \theta, \beta, k, D_T \propto \alpha^{n+e-1} \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} e^{-\beta \sum_{i=1}^n t_i^\alpha - f\alpha} \{1 - I_k(\beta T^\alpha)\}^{N'};$$

$$(v) k | \alpha, N', \theta, \beta, D_T \propto \frac{\beta^{kn}}{\{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} e^{-a_n k} I_{\{0,g\}}(k).$$

5.3 Simulated study

In this section we use a set of simulated data with a sample size of 300 and $\theta = 300$, $\beta = 0.02$, $\alpha = 1$ and $k = 1$.

For both sets 1 and 2 of prior specifications of model I we have problems of convergence of the parameters, therefore, we use a number of iterations equal to 900,000, burn-in of 50,000 and leaps of 500.

We also obtained posterior estimates of restricted models: model II, $k = 1$ and model III, $k = 1, \alpha = 1$, we used a number of iterations equal to 50,000, burn-in of

1,000 and leads of 10.

In model I, in which the set 1 of prior distributions of Achcar et al. (1998), are given by

- $N' \sim P \{ \theta [1 - I_k(\beta T^\alpha)] \}$;
- $\theta \sim \Gamma(0.001, 0.001)$;
- $\beta \sim \Gamma(0.001, 0.001)$;
- $\alpha \sim \frac{1}{\alpha}, \alpha > 0$;
- $k \sim \frac{1}{k}, k > 0$.

Let us call the model with this prior specification model I(1).

In Model I, using set 2 of prior specifications, we have that for the parameter α it has been assigned a specification of the prior distribution $\Gamma(0.001, 0.001)$ and for parameter k a study was conducted about sensitivity analysis to estimate this model, by proposing to it several prior distributions and different values for its hyperparameters, since when it is free we have difficulty of obtaining convergence of the parameters. The tested prior distributions for this parameter include: Uniform (0, T), Gamma(0.001,0.001), Gamma(100,100), besides other values for the hyperparameters of such distributions. The transformation $\log(k)$ was also tested, where $\log(K) \sim N(a_1, b_1)$, with various values of the hyperparameters a_1, b_1 , among them, $N(0;0.01)$, $N(10;10)$, $N(0;1)$. In most tests the convergence of the Markov chain simulations was not satisfactory. The best result for this model was obtained using a truncated exponential distribution (a_n, g) , where values of g from 3 to 20 were tested. The best result was for truncated exponential distribution with mean parameter 0.95 and truncated in 3. Let's call the model with such prior model I(2). Table 6.1 shows posterior summaries the estimates of models I, II and III.

Tabela 5.1: Posterior summaries of parameters for models I, II and III. True model: $\theta = 300$, $\beta = 0.02$, $\alpha = 1$ and $k = 1$. Sample size of 300. Priors: for parameters θ , β and α are the ones in section (5.2). Priors for the parameter in model I: (1) non-informative prior, proposed in the article of Achcar et al. (1998), (2) exponential prior distribution with mean parameter 0.95 and truncated in 3. For model I, we have: 1st row of each parameter corresponds to an estimate of chain 1, starting from a given value; 2nd row we have estimates of chain 2, starting from the true values.

Models and Priors	Parameters	Priors		Posterior estimates			
		Mean	S.D.	Mean	Median	S.D.	IC(95%)
I (1)	θ	1	31.62	313.15	312.97	18.63	(282.44 ; 344.89)
		1	31.62	313.49	313.04	18.35	(284.53 ; 343.25)
	β	1	31.62	0.220*	0.127*	0.224*	(0.013* ; 0.6855*)
		1	31.62	0.334*	0.084*	0.525*	(0.009* ; 1.502*)
	α	-	-	0.709*	0.687*	0.196*	(0.444* ; 1.070*)
		-	-	0.744*	0.757*	0.255*	(0.348* ; 1.127*)
	k	-	-	2.026*	1.787*	0.889*	(0.910* ; 3.6825*)
		-	-	2.202*	1.559*	1.5665*	(0.819* ; 5.588*)
I(2)	θ	1	31.62	310.13	309.60	17.40	(282.58 ; 339.62)
		1	31.62	310.39	310.35	17.63	(281.83 ; 339.85)
	β	1	31.62	0.036	0.023	0.041	(0.0030 ; 0.113)
		1	31.62	0.043	0.024	0.054	(0.0026 ; 0.166)
	α	1	10	0.984	0.967	0.1915	(0.696 ; 1.333)
		1	10	0.983	0.962	0.2155	(0.644 ; 1.355)
	k	0.95	0.57	1.105	1.058	0.336	(0.653 ; 1.736)
		0.95	0.57	1.133	1.058	0.403	(0.627 ; 1.971)
II	θ	1	31.62	310.190	310.04	17.87	(281.45 ; 340.15)
	β	1	31.62	0.019	0.019	0.004	(0.014 ; 0.027)
	α	-	-	0.999	0.999	0.049	(0.917 ; 1.078)
III	θ	1	31.62	310.09	309.44	17.38	(282.47 ; 339.11)
	β	1	31.62	0.019	0.019	0.001	(0.017 ; 0.021)

* there is no convergence for these parameters.

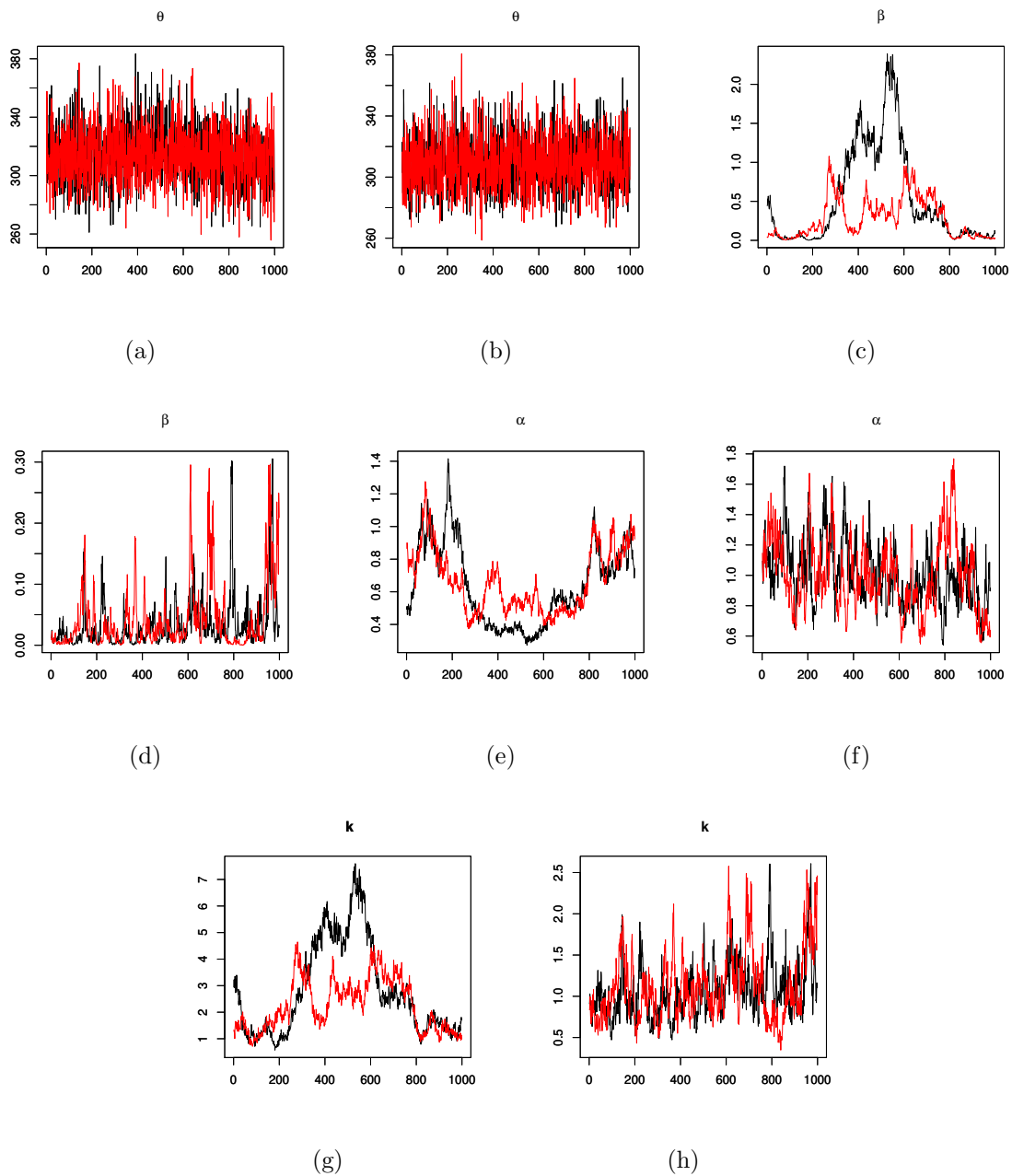


Figure 5.1: Convergence of chains 1 and 2, for the parameters of models: model I(1), in which graphs correspond to the letters: (a), (c), (e) and (g); model I(2), in which graphs correspond to the letters: (b), (d), (f) and (h).

For model I(1) which uses a non-informative prior distribution specification for parameter k the graphs of the MCMC chains are shown in Figure 5.1 (letters (a), (c), (e) and (g)). Only for parameter θ (letter (a)) appears to have clear convergence, while for parameter β (letter (c)) there is clear non convergence, and possible non convergence for parameter k (letter (g)). Figure 5.2 (letters (a), (c), (e) and (g)) presents the posterior distributions of the two chains. Only for parameter θ (letter (a)) there is indication of convergence. These are confirmed by the Gelman-Rubin's statistics which are equal to 1.25 and 1.15, for the parameters β and k respectively. So we did not get the convergence of the MCMC chains with these prior distributions.

In model I(2) we use the proposed prior distribution of the parameter k , an exponential distribution with mean parameter 0.95 and truncated in 3. The chains are presented in Figure 5.1 (letters (b), (d), (f) and (h)) and are indication of convergence. Figure 5.2 (letters (b), (d), (f) and (h)) shows the posterior distribution of parameters for chains 1 and 2, and there is also indication of convergence. The convergence is confirmed by the Gelman-Rubin's statistics, with the largest value equal to 1.01.

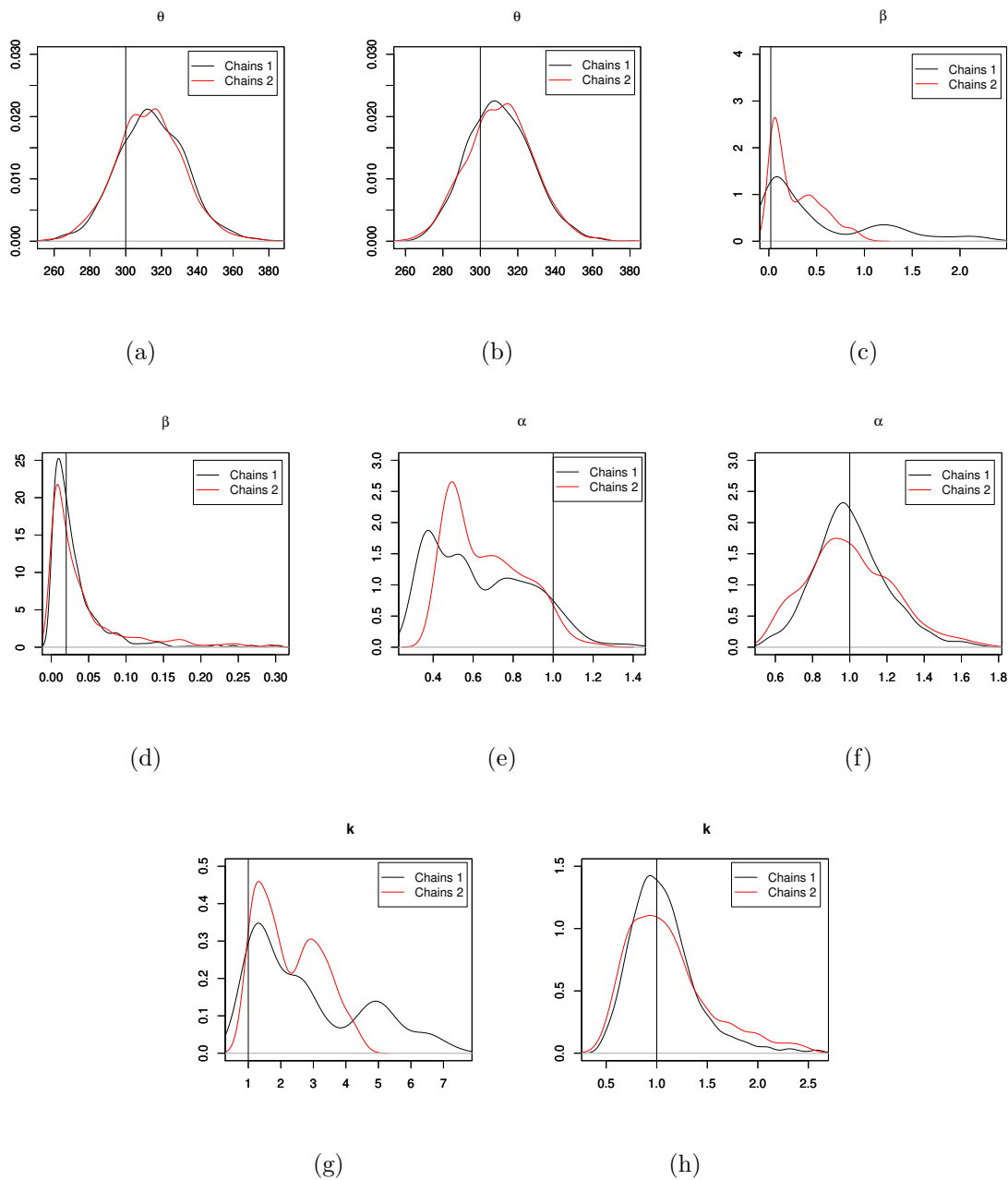


Figure 5.2: Posterior distribution of parameters concerning chains 1 and 2, of model I. Where graphs (a), (c), (e) and (g), correspond to the first set of priors and graphs (b), (d), (f) and (h) correspond to the second set of priors. The true value of each parameter is represented by the vertical line.

The main conclusion reached in this sensitivity analysis is that in the case of the proposed model 2, this is close to the process that generates the actual data, in other words, it is possible to have a good approximation of the posterior distribution of parameters. Even using little informative prior distributions, model I(2) can estimate the true value of the parameters. Not so with model I(1).

5.4 An application with pollution data

This section applies models I(1) and I(2) to fit data corresponding to the maximum daily mean measurements of ozone gas, based on data from the northeast (NE) region of Mexico City, with a sample of 981 observations, which correspond to times when a certain threshold established for air quality standard is violated by time T . These data are available at www.sma.df.gob.mx/simat/, which correspond to eighteen years of observations (1990 to 2008) (Achcar et al. 2009) the period 1 January 1990 to December 31, 2008.

For these models and the assumed data set we have used 700,000 iterations and a burn-in of 15,000, with leads of size 100.

Table 5.2 shows the posterior estimates of parameters for both models 1 and 2.

In Figure 5.3, we show the convergence of chains 1 and 2 respectively, for the parameters of models. Where the graphs that refer to letters: (a), (c), (e) and (g), correspond to model I(1), and graphs that refer to letters: (b), (d), (f) and (h) correspond to model I(2).

Tabela 5.2: Posterior estimates of parameters, using a sample of real data corresponding to the NE region, referring to Mexico City. Priors: for parameters θ , β and α are the ones in section (5.2). Priors for parameter k : (1) non-informative prior, proposed in the article of Achcar et al. (1998); (2) exponential prior distribution with mean parameter 0.99 and truncated in 6. In the 1st row, for each parameter, we present the estimates corresponding to chain 1, in the 2nd row, we present estimates corresponding to chain 2.

Priors for k		Priors		Posterior estimates				
		Mean	S.D.	Mean	Median	S.D.	IC(95%)	
1	θ	1	31.62	995.64	995.17	32.50	(942.78 ; 1049.44)	
		1	31.62	999.310	998.33	34.84	(944.71 ; 1057.79)	
	β	1	31.62	3.7x10 ⁻⁴ *	0.6x10 ⁻⁵ *	2.4x10 ⁻³ *	(0.8x10 ⁻⁶ * ; 1.3x10 ⁻⁴ *)	
		1	31.62	2.9x10 ⁻³ *	1.1x10 ⁻⁵ *	1.2x10 ⁻² *	(2.7x10 ⁻⁶ * ; 2.5x10 ⁻² *)	
	α	-	-	1.530*	1.569*	0.235*	(0.966* ; 1.808*)	
		-	-	1.425*	1.507*	0.269*	(0.663* ; 1.663*)	
	k	-	-	1.114*	1.002*	0.409*	(0.817* ; 2.007*)	
		-	-	1.332*	1.067*	0.823*	(0.915* ; 3.718*)	
	2	θ	1	31.62	1043.54	1043.38	35.39	(986.58 ; 1102.41)
			1	31.62	1044.71	1044.15	35.79	(985.53 ; 1103.67)
β		0,1	0,001	0.055	0.052	0.022	(0.025 ; 0.095)	
		0,1	0,001	0.056	0.053	0.021	(0.026 ; 0.094)	
α		1	10	0.592	0.589	0.042	(0.530 ; 0.662)	
		1	10	0.588	0.586	0.038	(0.529 ; 0.656)	
k		0.99	0.95	4.457	4.445	0.512	(3.660 ; 5.328)	
		0.99	0.95	4.486	4.467	0.491	(3.697 ; 5.324)	

* there is no convergence for these parameters.

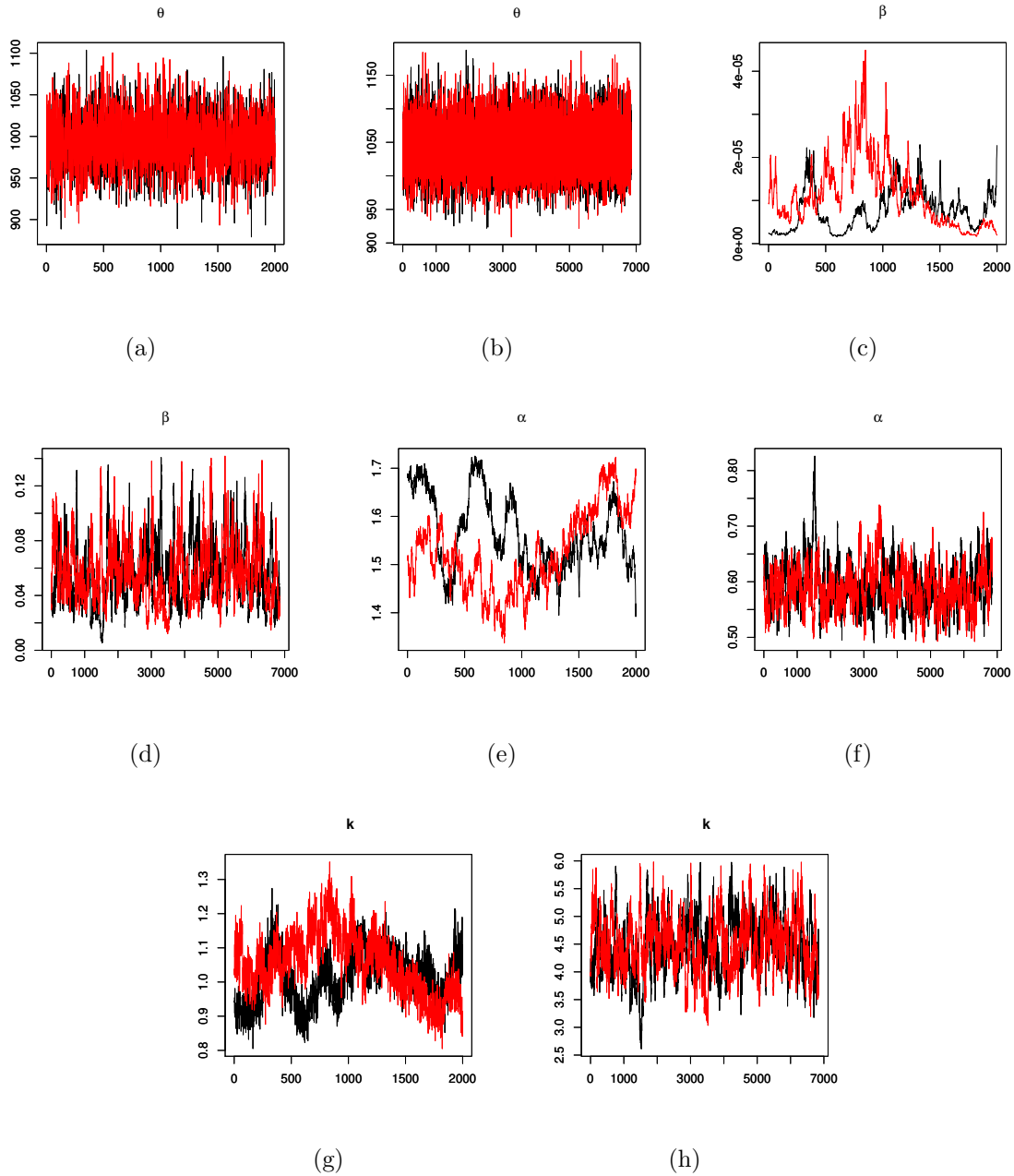


Figure 5.3: Convergence of chains 1 and 2, for the parameters of models: model I(1), in which graphs correspond to the letters: (a), (c), (e) and (g); model I(2), in which graphs correspond to the letters: (b), (d), (f) and (h).

We can observe in Figure 5.3, that the parameters that refer to letters: (a), (c), (e) and (g), which correspond to model I(1), in which we have used a non-informative prior distribution specification for parameter k , only for the parameter θ appears to have convergence, the same corresponds to letter (a), the same does not occur to (c), (e) and (g), ie, in this case we did not observe the convergence for the other parameters. In model I(2), the parameters that refer to letters: (b), (d), (f) and (h), appear to have convergence, which was proposed for parameter k an exponential priori distribution with mean parameter 0.99 and truncated in 6.

The largest value of the Gelman-Rubin's statistics for the parameters of the general model I(2) is equal to 1.00 indicating convergence of the MCMC chains. In the model I(1) the Gelman-Rubin's statistics are equal to 1.22, 1.23 and 1.19, for the parameters α , β and k respectively. So we did not get the convergence of MCMC chains using the non informative prior for parameter k .

In Figure 5.4, we have the posterior distribution of parameters for chains 1 and 2 for models I(1) and I(2). The graphs that refer to letters: (a), (c), (e) and (g), correspond to the first set of priors; and the graphs that refer to letters: (b), (d), (f) and (h) correspond to the second set of priors. The results confirm the convergence in the model I(2) and non convergenc in the model I(1).

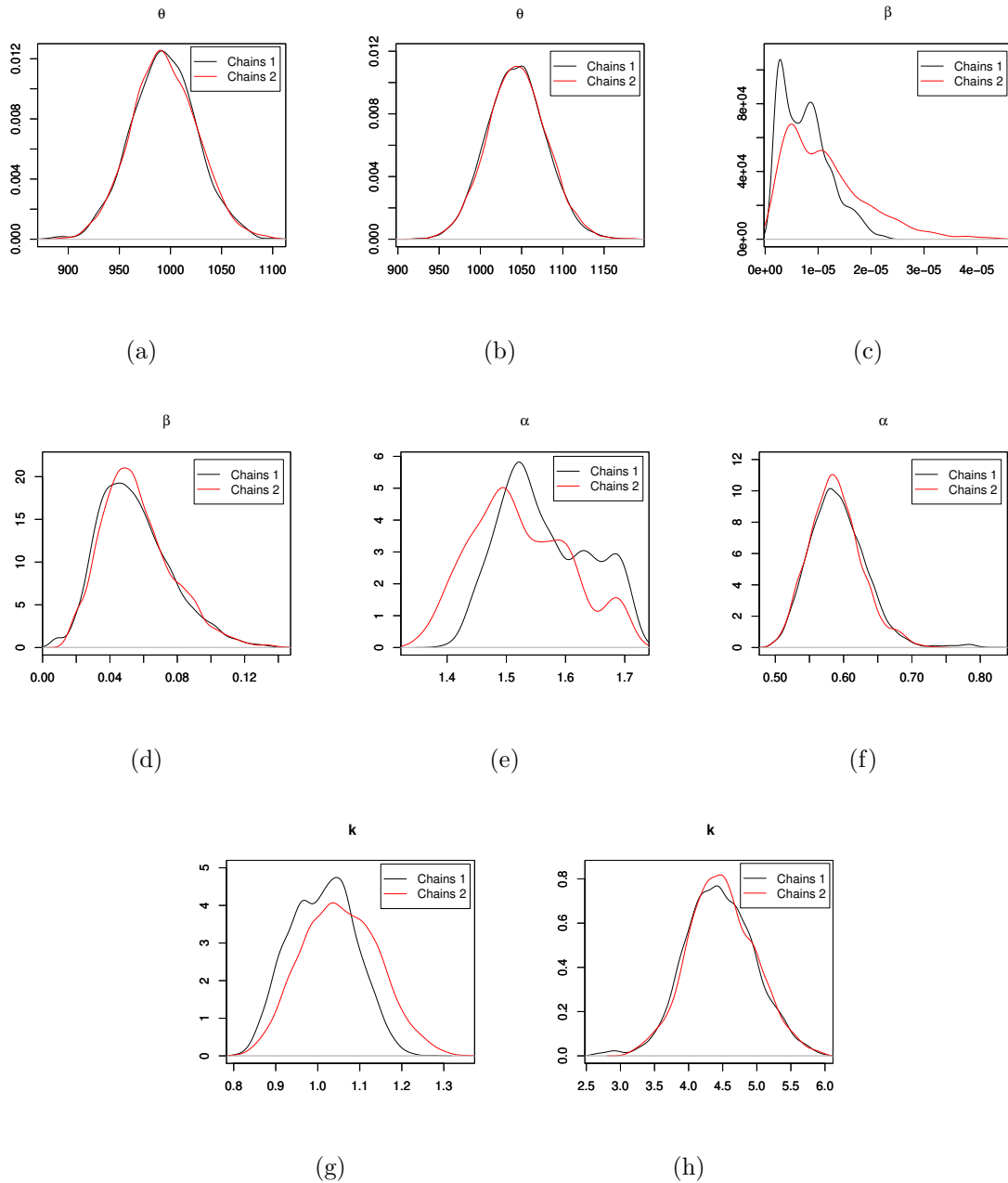


Figure 5.4: Posterior distribution of parameters for chains 1 and 2, respectively for model I. Where graphs (a), (c), (e) and (g), correspond to the first set of priors; and graphs (b), (d), (f) and (h) correspond to the second set of priors.

5.5 Conclusions

In this article we proposed a sensitivity analysis with various specifications of prior distributions for the model previously introduced by Achcar et al. (1998), which was developed under the Bayesian approach. We have hereby conducted a study of the effect of prior distributions on convergence and accuracy of results, thus being able to propose a prior distribution for parameter k , and when this parameter is estimated we have more difficult it is then to achieve the convergence of parameters. After trying several prior distribution specifications for this parameter, it was possible to propose a prior distribution that would considerably improve the convergence of the chains. Such improvements can be noted in the credible intervals, in which the range of the interval is smaller using a truncated exponential prior distribution; we may also observed such best results through graphical analysis, where when using a truncated exponential prior distribution for parameter k , we obtain the convergence of the chains.

Referências Bibliográficas

Achcar, J. A.; Dey, K.D. and Niverthi, M. –*A Bayesian approach using nonhomogeneous poisson process for software reliability models*, In: A.S.Basu; S.K.Basu; S.Mukhopadhyay, (Org.). *Frontiers in Reliability*. ed.1 River Edge, NJ: World Scientific, v. 4, p. 1–18, 1998.

Achcar, J. A.; Fernández-Bremauntz, A. A.; Rodrigues, E.R. E Tzintzun, G. –*Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model*, *Environmetrics*, v. 19, Issue 5, p. 469–485, 2008.

Achcar, J. A.; Rodrigues, E.R.; Paulino, C. D. and Soares, P. –*Non-homogeneous Poisson models with a change-point: an application to ozone peaks in Mexico city*, *Environmental and Ecological Statistics*, v. 17, p. 521–541, 2010.

Casella, G. and Berger, R. L. –*Statistical Inference*, Second Edition, 2001.

Cox, D.R. and Lewis, P.A. –*Statistical Analysis of Series of Events*, Methuen, 1966.

Iman, A. Inference and monitoring convergence. *Markov Chain Monte Carlo in Practice* (eds. W. R. Gilks, S. Richardson e D. J. Spiegelhalter), London: Chapman and Hall, p. 131–143, 1996.

Iman, A. and Rubin, D. R. A single series from the gibbs sampler provides a false sense of security. *Bayesian Statistics 4* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith), Oxford: Oxford University Press, p. 625–631, 1992.

- Goel, A.L. and Okumoto, K. –*An Analysis Of Recurrent Software Errors In A Real-Time Control System*, In Proceedings of ACM Conference, Washington, D.C., USA, p. 496–501, 1979.
- Goel, A.L. –*A guidebook for software reliability assessment*, Technical Report, RADC-TR-83-176, 1983.
- Hastings, W. K. –*Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, Biometrika, v. 57, p. 97–109, 1970.
- Kou, L, and Yang, T.Y. –*Bayesian computation for nonhomogeneous Poisson process in software reliability*, To appear in Journal of the American Statistical Association, 1995.
- Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. –*Equations of State Calculations by Fast Computing Machines*, Journal of Chemical Physics, v. 21, p. 1087–1092, 1953.
- Ohba, M.; Yamada, S.; Takeda, K. and Osaki, S. –*S-shaped software reliability growth curve: how good is it?*, COMPSAC’82, p. 38–44, 1982.
- Yamada, S.; Ohba, M. and Osaki, S. –*S-Shaped reliability growth modeling for software error detection*, IEEE Transactions on Reliability, v. 32, p. 475–478, 1983.
- Yang, T.Y. –*Computacional approaches to bayesian inference for software reliability*, Ph.D. Thesis, University of Connecticut, Storrs, 1994.
- Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P. and Van der Linde, A. –*Bayesian measures of model complexity and fit (with discussion)*, Journal of the Royal Statistical Society, B(64), p. 583–639, 2002.

Capítulo 6

Non-homogeneous Poisson process in the presence of one or more change-points: an application to air pollution data

Lorena Vicini, Luiz Koodi Hotta and Jorge Alberto Achcar

Abstract. We consider the problem of modeling the number of times that an air quality standard is exceeded in a certain period of time. We assume that the number of times the threshold is exceeded takes place according to a non-homogeneous Poisson process (NHPP) with the mean function modeled by the generalized gamma distribution. We consider models with and without change-points. When the presence of change-points is assumed, we have none, one, two or three change-points, depending on the data set. We use the Bayesian approach, where the posterior summaries of interest are obtained using standard Markov Chain Monte Carlo (MCMC) methods. We also discuss the use of different prior distributions for the parameters of the models, with an analysis of the convergence of the Gibbs sampling algorithm and sensitivity for the choice of different priors. To illustrate the proposed method we consider simulated data and a pollution data set from a region of Mexico City.

6.1 Introduction

One problem that has affected many regions around the world is air pollution. In some places, such as in big cities and industrial regions, there is a higher concentration of pollution. However, due to wind, the air pollution can spread to other regions .

Air pollution has become a public health problem, since an increase in pollution can cause serious public health problems, such as diseases related to respiratory and cardiovascular systems; these have been highlighted in many health studies (see, for example, Braga et al., 2002 ; Gouveia et al., 2006). Air pollution is characterized by the presence of toxic gases and liquid or solid particles in the air. An important example of a pollutant is ozone, because when its concentration remains above a threshold level for a certain period of time, individuals exposed to it can suffer serious health problems (see, for example, Air Resource Board (ARB), 2005).

In this paper we analyze a series of data for ozone (O_3), which is a gas composed of three oxygen atoms and formed by chemical reactions between nitrogen oxides (NOx) and volatile organic compounds ($VOC's$) in the presence of sunlight. Ozone has the same chemical structure miles above the earth or at ground level and it can be "good" or "bad" depending on its location in the atmosphere. In the lower atmosphere, the tropospheric ozone is considered "bad".

As pollution levels have increased at an alarming rate in recent years, exceeding the limits for acceptable standard of air quality on certain days, studies related to the problem are gaining prominence around the world. As a direct result, new models and statistical methods have been developed to analyze air pollution data (see, for example, Achcar et al., 2011). In some studies the authors have used classical methods of statistical inference, such as the use of generalized linear models and correlation coefficients. Other studies have used the Bayesian approach.

In relation to pollution by ozone gas, the literature contains several studies (see, for example, Wilson et al., 1980; Loomis et al., 1996; Galizia and Kinney, 1999; Bell et al., 2004; Gauderman et al., 2004; ARB, 2005; Bell et al., 2005; Bell et al., 2007 ; and

Achcar et al., 2009).

Other studies are also related to diseases caused by an increase in the level of air pollution (see, for example, Martins et al., 2002; Farhat et al., 2005).

In this paper, we consider how to model the data by a non-homogeneous Poisson process (NHPP) with the rate function modeled by the generalized gamma distribution. The model is used to analyze the daily data set collected by the monitoring network of the Metropolitan Area of Mexico City. The set contains 18 years of daily average ozone measurements in the period from 1 January 1990 to 31 December 2008. The Metropolitan Area of Mexico City is divided into five regions, corresponding to the Center (CE), Northwest (NW), Northeast (NE), Southeast (SE) and Southwest (SW).

This paper considers the modeling of the data in the presence or not of one or more change-points. We get posterior summaries of interest using standard MCMC methods, in special, the Gibbs sampling algorithm or the Metropolis-Hastings algorithm. Also it is discussed the sensitivity of the choice of different prior distributions for the parameters of the model and their effect in the convergence of the simulation algorithm. The studies are illustrated with simulated data and a pollution real data set.

This article is organized as follows: in Section 6.2, the model is presented; Section 6.3 presents the Bayesian formulation, first without taking into account change-points, which are incorporated later; examples considering simulated data set are given in Section 6.4; in Section 6.5, the proposed models are applied to a data set collected by the monitoring network of the Metropolitan Area of Mexico City; section 6.6 concludes with final remarks and discussions of the results.

6.2 Description of the model

The NHPP model has been applied in various areas. The problem can be described as follows. Let $T > 0$ be a real number and $M = \{M(t) : t \in (0, T]\}$, where the random variable $M(t)$ represents the cumulative number of events in the time interval $[0, t)$ for $t \geq 0$. In the NHPP model the random variable $M(t)$ has a Poisson distribution

with mean $m(t)$. One can also characterize the distribution by the intensity function $\lambda(t) = \frac{dm(t)}{dt}$.

If $\lambda(t)$ is a constant, so that $m(t)$ is linear, then $M(t)$ is called a homogeneous Poisson process; otherwise the process is called a non-homogeneous Poisson process.

Different choices for the function $m(t)$ are considered in the literature, especially in software reliability modeling (see, for example, Achcar et al., 1998). Goel and Okumoto (1979) stated that the expected number of software failures for time t is given by the mean value function $m(t)$, which is non-decreasing and bounded above. Specifically, they considered the mean function

$$m_1(t) = \theta (1 - e^{-\beta t}), \quad (6.1)$$

where, in our case, θ represents the expected maximum number of days in which the air quality standard is violated by a particular pollutant and β is considered to be the rate at which events occur. By (6.1), we have that In Goel (1983) a generalization of the model (6.1) was proposed, given by the following intensity function

$$\lambda_2(t) = \theta \beta \alpha t^{\alpha-1} e^{-\beta t^\alpha}, \quad (6.2)$$

The corresponding mean function is

$$m_2(t) = \theta (1 - e^{-\beta t^\alpha}). \quad (6.3)$$

Note that (6.1) and (6.3) can be written as special cases of the general form where the mean value function is given as

$$m(t) = \theta F(t), \quad (6.4)$$

where $F(t)$ is a distribution function. On the other hand, for any distribution function $F(t)$ we have a valid model.

A widely used distribution function, given its high flexibility, is the generalized gamma distribution. In this case the mean function is

$$m_{GG}(t) = \theta I_k(\beta t^\alpha), \quad (6.5)$$

where $I_k(s)$ is the integral of the gamma function given by

$$I_k(s) = \frac{1}{\Gamma(k)} \int_0^s x^{k-1} e^{-x} dx. \quad (6.6)$$

From (6.6), we obtain the intensity function given by

$$\lambda_{GG}(s) = m'_{GG}(t) = \frac{1}{\Gamma(k)} \theta \beta^k \alpha t^{\alpha k - 1} e^{-\beta t^\alpha}. \quad (6.7)$$

This model is called generalized gamma. When k is an integer we can write $F(t)$ as

$$F(t) = 1 - e^{-\beta t^\alpha} \sum_{j=0}^{k-1} \frac{(\beta t^\alpha)^j}{j!}. \quad (6.8)$$

The three models used in this article are described as follows:

- a) Model I: Here all parameters are unknown, $\lambda_{GG}(t)$ is given by (6.7).
- b) Model II: With $k = 1$, $\lambda_{GG}(t)$ is given by $\lambda_2(t)$ given in (6.2) Goel (1983).
- c) Model III: With $k = 1$ and $\alpha = 1$, $\lambda_{GG}(t)$ reduces to $\lambda_1(t)$ Goel and Okumoto (1979).

In the Bayesian analysis of these models, we may have some difficulties in obtaining the Bayesian inferences using MCMC simulation methods. The difficulties are mainly related to the convergence of the MCMC chains. For this reason we will study the effect of different prior distributions on the performance of sample simulation algorithms of the posterior distribution of interest.

The main idea of this paper is to model the number of times that an air quality standard is exceeded in a period of time, under the Bayesian approach. To do this we explore the generalized intensity functions that can best fit the data set. In particular, we fit the generalized intensity function with and without the presence of change-

points. We also study several prior distributions for the parameters because of the convergence problem mentioned above, especially when all the parameters are estimated simultaneously. For this reason we perform sensitivity analysis in relation to prior distributions. The biggest concern is with the parameter k , a parameter of the gamma distribution.

The convergence of convergence of the MCMC algorithm was analyzed by graphical method and by the Gelman-Rubin statistics. The Gelman-Rubin statistics (1992) relies on parallel chains to test whether they all converge to the same posterior distribution. Brooks and Gelman (1997) suggested the introduction of a correction factor. This statistics is evaluated using the coda package in R. This statistics is larger or equal to 1.0. The closer this statistics is to 1.0, more evidence we have that the chain is near convergence. The limit value of 1.2 is sometimes used as a guideline for “approximate convergence” (Gelman, 1996).

6.3 Bayesian inference

Denote by $D_T = \{n; t_1, t_2, \dots, t_n; T\}$, the data set, where n is the number of events observed such that $0 \leq t_1 < t_2 < \dots < t_n \leq T$, and where t_i are the times of the events observed during the period of time $(0, T]$.

We consider that the parameters θ, β, α and k are unknown and will be estimated. In the Bayesian framework, for each parameter, we must select prior distributions which describe the uncertainty about them.

In this article we consider the presence or absence of change-points to NHPP. We have two different forms for the likelihood function of the model, one for each formulation. First, we define the notation and expressions for the case without change-points, and then we do the same for the case when there are one or more change-points.

6.3.1 Models without change-points

For Model I, the likelihood function for the vector $\Theta = (\theta, \beta, \alpha, k)$, considering T as the truncation time of the truncated model (see, for example, Cox and Lewis, 1966), is given by

$$L(\Theta | D_T) = \left(\prod_{i=1}^n \lambda(t_i) \right) \exp(-m(T)). \quad (6.9)$$

In some cases it is advisable to enter a latent variable, as this may serve as a computational aid. We consider the introduction of the latent variable N' which has Poisson distribution with parameter $\theta[1 - F(T | \beta)]$ (see, for example, Achcar et al., 1998).

Considering the generalized gamma distribution model given in (6.5), the likelihood function for the parameters $\Theta = (\theta, \beta, \alpha, k)$ is expressed as

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \theta \frac{\beta^k}{\Gamma(k)} \alpha t_i^{\alpha k - 1} e^{-\beta t_i^\alpha} \right\} \exp \left\{ - \int_0^T \theta \frac{\beta^k}{\Gamma(k)} \alpha u^{\alpha k - 1} e^{-\beta u^\alpha} du \right\},$$

such that

$$L(\Theta | D_T) = \frac{\theta^n \alpha^n \beta^{kn}}{\{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k - 1} \right\} \exp \left\{ -\beta \sum_{i=1}^n t_i^\alpha - \theta I_k(\beta T^\alpha) \right\}. \quad (6.10)$$

Prior distributions

The likelihoods for Models II and III are particular cases of the likelihood for Model I.

This work used two sets of a priori distributions for Model I; the first set (1) of a

priori distributions are the distributions suggested by Achcar et al. (1998)

- (i) $N' \sim \text{Poisson} \{ \theta [1 - I_k(\beta T^\alpha)] \}$;
- (ii) $\theta \sim \text{Gamma}(a, b)$; known a and b ;
- (iii) $\beta \sim \text{Gamma}(c, d)$; known c and d ;
- (iv) $\alpha \sim \pi_1(\alpha)$ where $\pi_1(\alpha) \propto \frac{1}{\alpha}$ for $\alpha(\alpha > 0)$;
- (v) $k \sim \pi_2(k)$ where $\pi_2(k) \propto \frac{1}{k}$ for $k(k > 0)$.

In the second set (2), the prior distributions for α and k are modified as

- (iv) $\alpha \sim \text{Gamma}(e, f)$; known e and f ;
- (v) $k \sim \text{truncated exponential distribution}(a_n, g)$; known a_n and g .

$P(\lambda)$ denotes the Poisson distribution with parameter λ , a_n , g are the hyperparameters of the truncated exponential distribution, and a, b, c, d, e and f are known hyperparameters of the gamma distributions where $\text{Gamma}(a, b)$ denotes a gamma distribution with expected value $\frac{a}{b}$ and variance $\frac{a}{b^2}$.

We assume independence of the prior distributions of parameters. The values of the hyperparameters are given in the applications. As we may have convergence problems in this model, in the applications we study the effects of other prior distributions on some posterior distributions of parameters. These prior distributions are given in Section 6.4.

In the restricted Models II and III we use the same prior distributions for the free parameters.

Posterior distributions

The inference will be conducted based on information supplied by the posterior distribution of the parameters. Assuming independence of the prior distributions, the likelihood function for the Poisson processes, both for the homogeneous case and for

the non-homogeneous case, are given as

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \lambda(t_i) \exp \left(- \int_{t_{i-1}}^{t_i} \lambda(x) dx \right) \right\} \exp \left\{ - \int_{t_n}^T \lambda(u) du \right\},$$

that is,

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \lambda(t_i) \right\} \exp \left\{ - \int_0^T \lambda(u) du \right\}.$$

Thus, we have

$$L(\Theta | D_T) = \left\{ \prod_{i=1}^n \lambda(t_i) \right\} \exp \{-m(T)\}. \quad (6.11)$$

Initially, we assume the prior distribution for k proposed by Achcar et al.(1998).

The posterior distribution is given by

$$P(N', \alpha, \beta, \theta / D_T) \propto \frac{\theta^{N'+n+a-1} \alpha^n \beta^{kn+c-1}}{N'! \{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k-1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} \\ \times e^{-(b+1)\theta - (d + \sum_{i=1}^n t_i^\alpha) \beta} \pi_1(\alpha) \pi_2(k). \quad (6.12)$$

Since the posterior distribution in (6.12) has no closed form, we resort to MCMC methods to simulate samples of the joint posterior distribution. The algorithm used to obtain posterior distribution samples in (6.12) is given as follows. For the subset of parameters whose full conditional posterior distributions is known, we sample directly from it, using the Gibbs Sampling algorithm and for: and for the subset of parameters in which the conditional densities are not known, the samples are simulated using the steps of the Metropolis-Hastings algorithm (see Metropolis et al., 1953; Hastings, 1970). Recommended references for the review of MCMC methods are given in Casella and Berger (2001). The required full conditional posterior distributions needed for the MCMC algorithms are given by,

- (i) $N' | \theta, \alpha, \beta, k, D_T \sim Poisson[\theta(1 - I_k(\beta T^\alpha))];$
- (ii) $\theta | N', \alpha, \beta, k, D_T \sim Gamma[a + n + N', b + 1];$

- (iii) $\beta \mid N', \alpha, \theta, k, D_T \propto \beta^{kn+c-1} e^{-\beta[\sum_{i=1}^n x_i^\alpha + d]} \{1 - I_k(\beta T^\alpha)\}^{N'}$;
 (iv) $\alpha \mid N', \theta, \beta, k, D_T \propto \alpha^n \left\{ \prod_{i=1}^n t_i^{\alpha k-1} \right\} e^{-\beta \sum_{i=1}^n t_i^\alpha} \{1 - I_k(\beta T^\alpha)\}^{N'} \pi_1(\alpha)$;
 (v) $k \mid \alpha, N', \theta, \beta, D_T \propto \frac{\beta^{kn}}{\{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k-1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} \pi_2(k)$.

The parameter θ s have closed form distributions, so their posterior distributions are obtained through the Gibbs sampling method. For the other parameters we use the Metropolis-Hastings algorithm.

When we adopt the prior specification of the second set, proposed in Model I, we have that the posterior distribution is given by

$$P(N', \alpha, \beta, \theta / D_T) \propto \frac{\theta^{N'+n+a-1} \alpha^{n+e-1} \beta^{kn+c-1}}{N'! \{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k-1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} \\ \times e^{-(b+1)\theta - (d + \sum_{i=1}^n t_i^\alpha)\beta - f\alpha - a_n k} I_{\{0,g\}}(k). \quad (6.13)$$

and the only changes in the full conditional posterior distributions occur in (iv) and (v), which are replaced by

- (iv) $\alpha \mid N', \theta, \beta, k, D_T \propto \alpha^{n+e-1} \left\{ \prod_{i=1}^n t_i^{\alpha k-1} \right\} e^{-\beta \sum_{i=1}^n t_i^\alpha - f\alpha} \{1 - I_k(\beta T^\alpha)\}^{N'}$;
 (v) $k \mid \alpha, N', \theta, \beta, D_T \propto \frac{\beta^{kn}}{\{\Gamma(k)\}^n} \left\{ \prod_{i=1}^n t_i^{\alpha k-1} \right\} \{1 - I_k(\beta T^\alpha)\}^{N'} e^{-a_n k} I_{\{0,g\}}(k)$.

6.3.2 Models with change-points

The ozone pollution often changes during the time interval $(0, T]$, due to some type of intervention or change. In special, some political decisions by public authorities could implies in a decreasing or an increasing in ozone gas emission and, therefore, a similar effect can occur in the daily measurements of ozone.

In this case we can have J change-points, which we denote by τ_j , $j = 1, \dots, J$, and in each interval we use the generalized gamma model, the NHPP model, presented in Section 6.2.

We assume here that all the parameters are unknown: $(\theta_i, \alpha_i, \beta_i, k_i)$ and τ_j , where $i = 1, \dots, J + 1$ and $j = 1, \dots, J$. They must be estimated, and the change-points can occur at any time $\tau_0, \tau_1, \dots, \tau_i$, where $\tau_0 = 0$. We initially include a single change-point

in the model; a second and third change-point will be included later.

In this case, we have that the rate function of the NHPP process is of the form (Achcar et al., 2009)

$$\lambda(t | \Theta) = \begin{cases} \lambda(t | \Theta_1), & \text{se } 0 \leq t < \tau_1 \\ \lambda(t | \Theta_j), & \text{se } \tau_{j-1} \leq t < \tau_j, j = 2, 3, \dots, J, \\ \lambda(t | \Theta_{J+1}), & \text{se } \tau_J \leq t \leq T, \end{cases} \quad (6.14)$$

where $\lambda(t | \Theta_j), j = 1, 2, \dots, J + 1$ intensity functions are related to equation (12) and $\Theta_j, j = 1, 2, \dots, J + 1$ are the parameters associated with the NHPP process in each interval limited by the change-points.

For the generalized gamma distribution the rate function is of the form

$$\lambda(t | \Theta) = \frac{1}{\Gamma(k_j)} \theta_j \beta_j^{k_j} \alpha_j t^{\alpha_j k_j - 1} e^{-\beta_j t^{\alpha_j}}.$$

The mean function $m(t | \Theta_j), j = 1, 2, \dots, J + 1$, as given in Achcar et al. (2009) is of the form

$$m(t | \Theta) = \begin{cases} m(t | \Theta_1), & \text{se } 0 \leq t < \tau_1 \\ m(\tau_1 | \Theta_1) + m(t | \Theta_2) - m(\tau_1 | \Theta_2), & \text{se } \tau_1 \leq t < \tau_2, \\ m(t | \Theta_{j+1}) - m(\tau_j | \Theta_{j+1}) \\ + \sum_{i=2}^j [m(\tau_i | \Theta_i) - m(\tau_{i-1} | \Theta_i)] + m(\tau_1 | \Theta_1), & \text{se } \tau_j \leq t < \tau_{j+1}, \\ j = 2, 3, \dots, J, \end{cases} \quad (6.15)$$

where $\tau_{J+1} = T$. For the generalized gamma distribution the mean function is given by

$$m(t | \Theta) = \theta_j I_{k_j}(\beta_j t^{\alpha_j}),$$

in which

$$I_{k_j}(s) = \frac{1}{\Gamma(k_j)} \int_0^s x^{k_j-1} e^{-x} dx.$$

Let $\mathbf{w} = (\Theta_1, \Theta_2, \dots, \Theta_{J+1}; \tau_1, \tau_2, \dots, \tau_j)$ in which $\Theta_i = (\theta_i, \beta_i, \alpha_i, k_i)$ in which $i = 1, \dots, J + 1$ and $j = 1, \dots, J$. Having assumed an NHPP, the likelihood function is given by

$$L(w | D_T) \propto \prod_{i=1}^{N_{\tau_1}} \lambda(t_i | w_1) e^{-m(\tau_1 | w_1)} \left[\prod_{j=2}^J \left(\prod_{i=N_{\tau_{j-1}}+1}^{N_{\tau_j}} \lambda(t_i | w_j) e^{-[m(\tau_j | w_j) - m(\tau_{j-1} | w_j)]} \right) \right] \\ \times \prod_{i=N_{\tau_J}+1}^{N_T} \lambda(t_i | w_{J+1}) e^{-[m(T | w_{J+1}) - m(\tau_J | w_{J+1})]}. \quad (6.16)$$

In the case of the generalized gamma distribution the likelihood function is given by

$$L(w | D_T) \propto \prod_{i=1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)} \theta_i \beta_i^{k_i} \alpha_i d_i^{\alpha_i k_i - 1} e^{-\beta_i d_i^{\alpha_i}} e^{-\theta_i I_{k_i}(\beta_i \tau_j^{\alpha_i})} \prod_{i=N_{\tau_j}+1}^{N_T} \frac{1}{\Gamma(k_i)} \theta_i \beta_i^{k_i} \alpha_i d_i^{\alpha_i k_i - 1} \\ \times e^{-\beta_i d_i^{\alpha_i}} e^{-[\theta_i I_{k_i}(\beta_i T^{\alpha_i}) - \theta_i I_{k_i}(\beta_i \tau_j^{\alpha_i})]}.$$

Prior distributions for models with change-points

The prior distributions for the parameters are given as follows:

- (i) $\tau_j \sim Uniform(f_j, g_j), j = 1, 2;$
- (ii) $N'_i \sim Poisson[\theta_i(1 - I_{k_i}(\beta_i \tau_i^{\alpha_i}))], i = 1, 2, 3;$
- (iii) $\theta_i \sim Gamma(a_i, b_i), i = 1, 2, 3, a_i, b_i$ known;
- (iv) $\beta_i \sim Gamma(c_i, d_i), i = 1, 2, 3, c_i, d_i$ known;
- (v) $\alpha_i \sim Gamma(e_i, h_i), i = 1, 2, 3, e_i, h_i$ known;
- (vi) $k_i \sim$ truncated exponential $(m_i, u), i = 1, 2, 3$ and m_i, u known.

We also assume independence among the prior distributions.

Let N'_i be a latent variable that has Poisson distribution with parameters $P[\theta_i(1 - I_{k_i}(\beta_i \tau_j^{\alpha_i}))]$, and let $\Phi = (\mathbf{w}, N'_i)$. The inference is performed using information supplied by the posterior distribution of the parameters. Assuming independence among the prior distributions, in the presence of change-points the

posterior and the prior distributions and the likelihood function are related as follows,

$$P(\Phi | D_T) \propto L(D_T | \Phi)P(\Phi), \quad (6.17)$$

where $P(\Phi | D_T)$ is the joint posterior distribution of Φ conditional to data D_T ; $P(\Phi)$ corresponds to the prior distributions of parameters and $L(D_T | \Phi)$ is the likelihood function.

The full conditional posterior distributions needed for the Gibbs sampling algorithm are given by,

- $P(N'_i) \sim P[\theta_i(1 - I_{k_i}(\beta_i \tau_j^{\alpha_i}))]$;
- $P(\theta_i) \sim \text{Gamma}(N_{\tau_j} + N'_i + a_i - 1, 1 + b_i)$;
- $P(\beta_i) \propto \beta_i^{K_i N_{\tau_j} + c_i - 1} e^{-\beta_i(\sum_{i=1}^{N_{\tau_j}} t_i^{\alpha_i} + d_i)} [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N'_i}$;
- $P(\alpha_i) \propto \alpha_i^{N_{\tau_j} + e_i - 1} \left(\prod_{i=1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1} \right) e^{-\beta_i \sum_{i=1}^{N_{\tau_j}} t_i^{\alpha_i} - h_i \alpha_i} [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N'_i}$;
- $P(k_i) \propto \left(\prod_{i=1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)} \right) \beta_i^{k_i N_{\tau_j}} \left(\prod_{i=1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1} \right) e^{-m_i k_i} I_{(0,u)} k_i [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N'_i}$;
- $P(\tau_j) \propto \left(\prod_{i=1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)} \right) \theta_i^{N_{\tau_j}} \beta_i^{k_i N_{\tau_j}} \alpha_i^{N_{\tau_j}} \left(\prod_{i=1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1} \right) e^{-\beta_i \sum_{i=1}^{N_{\tau_j}} t_i^{\alpha_i}}$
 $\times \left(\prod_{i=N_{\tau_j}+1}^{N_{\tau_j}} \frac{1}{\Gamma(k_i)} \right) \theta_i^{N_{\tau_j} - N_{\tau_j - 1}} \beta_i^{k_i N_{\tau_j} - N_{\tau_j - 1}} \alpha_i^{N_{\tau_j} - N_{\tau_j - 1}} \left(\prod_{i=N_{\tau_j}+1}^{N_{\tau_j}} t_i^{\alpha_i k_i - 1} \right)$
 $\times e^{-\beta_i \sum_{i=N_{\tau_j}+1}^{N_{\tau_j}} t_i^{\alpha_i}} [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i})]^{N'_i} [1 - I_{k_i}(\beta_i \tau_j^{\alpha_i}) + I_{k_i}(\beta_i \tau_{j-1}^{\alpha_i})]^{N'_i}$.

The parameters θ_i have distributions in closed form, so their posterior distributions are obtained using Gibbs Sampling, and for the other parameters α_i, β_i, k_i and τ_j the samples are obtained using the Metropolis-Hastings algorithm.

The inclusion of change-points in the restricted Models II and III is carried out in the same way as the inclusion in Model I.

6.4 Application to simulated data

We consider models without change-points and with one and two change-points. These models will initially be tested with simulated data sets and later will be used to

model ozone pollution data from Mexico City based on when a threshold in the levels of ozone concentration is exceeded for a certain period of time. In both cases we will discuss the convergence of the chains and the adjustment of the function $m(t)$. In most cases the number of simulated data is not large because we want to test whether the method works in this situation. Thus, we do not expect to find small variances for the posterior distributions.

6.4.1 Simulated data without change-point

We simulated 300 observations with $\theta = 300$, $\beta = 0.02$, $\alpha = 1$ and $k = 1$, i.e. Model III without a change-point. Models I and II are also correct because they incorporate this model, but they are over-specified.

We obtained a summary of estimates of the posterior distributions of Model I based on 900,000 MCMC replications after a burn-in of 50,000 and jumps of 500. For the restricted Models II and III, summaries of the estimates of the posterior distributions were obtained using 100,000 iterations, a burn-in of 10,000 and jumps of 10.

As the analysis of the adjustment of the models without change-point was carried out in Vicini et al. (2011), we only present the fit of the three models and the summaries of the estimated posterior distributions.

Figure 6.1 shows the adjustment of Models I, II and III in the absence of a change-point.

Figure 6.1 presents the graphs of the theoretical function $m(t)$, the empirical estimation, and its estimation by Models I, II and III without a change-point. All the models fit this simulated data set well. We observe that the estimated curves are superimposed on each other and on the empirical curve.

The curves are evaluated at the parameter values given by the median of the posterior distribution, since these provide better results than using the mean values. From now on we will always use the medians.

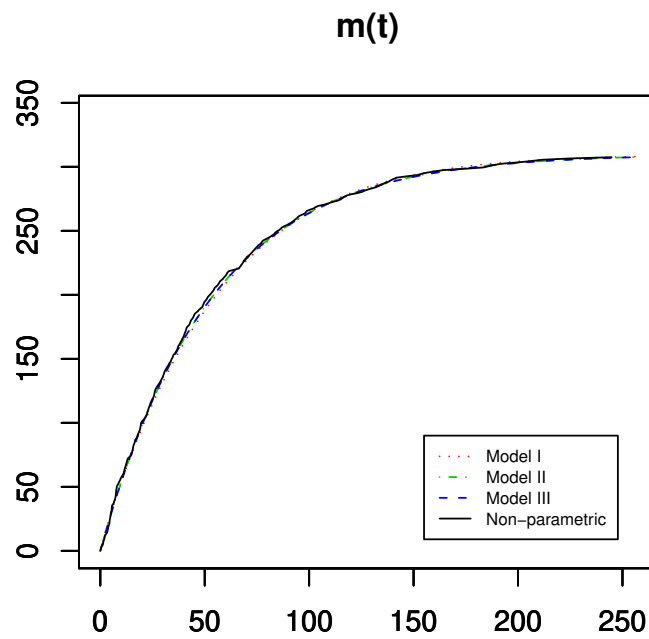


Figura 6.1: Simulation from the model without change-point with parameters $(k, \alpha) = (1, 1)$. Sample size equal to 300. Graphs of the theoretical function $m(t)$, the empirical estimation and its estimates by models without change-point: a) Model I, b) Model II c) Model III.

Table 6.1 presents the estimates of the mean, median and standard deviation of the parameters of Models I, II and III without change-points. The 95% credible intervals (CI) were obtained by taking the 2.5% and 97.5% percentiles. The parameter values were generally in their respective 95% CI. In this paper, we use the 2.5% and 97.5% percentiles for all the CI.

Tabela 6.1: Summary of posterior distributions of the parameters for Models I, II and III, with no change-points. Real model: $\theta = 300$, $\beta = 0.02$, $\alpha = 1$ and $k = 1$. Sample size of 300. Prior distributions for the parameters θ , β and α are presented in Section (6.3). The prior distributions for the parameter k in Model I is the exponential distribution with mean parameter equal to 0.95 and truncated at 3.

Model	parameter	mean	median	S.D.	(CI)2.5%	(CI)97.5%
I	θ	310.00	309.47	17.86	282.30	340.10
	α	1.234	1.219	0.174	0.976	1.543
	β	0.007687	0.005503	0.007301	0.000941	0.022115
	k	0.765	0.751	0.161	0.528	1.057
II	θ	310.28	309.89	17.63	282.27	339.70
	α	1.002	1.002	0.049	0.922	1.085
	β	0.019282	0.018933	0.004088	0.013193	0.026475
III	θ	310.03	309.82	17.77	281.62	339.49
	β	0.019115	0.019087	0.001214	0.017144	0.021113

We can observe in Table 6.1 that the mean and median estimates of the parameters of the subsequent Models I, II and III were satisfactory, because their values are very close to the true values of the parameters. We can also check that all CI contain the true value of the parameters. We observed good convergence of the simulation algorithm, as observed in the trace plots of the simulated samples. The largest value of the Gelman-Rubin's statistics is equal to 1.01 also indicating convergence of the MCMC chains. The results indicate that the information that $\alpha = k = 1$ is not relevant to the estimation of the parameter θ , as the estimate of the posterior distribution remains almost the same whether or not the constraint is taken into account.

6.4.2 Simulated data with one change-point

We simulated 250 observations with $\theta_1 = 300$, $\theta_2 = 330$, $\beta_1 = 0.0001$, $\beta_2 = 0.003$, $\alpha_1 = 2$, $\alpha_2 = 1$, $\tau = 90$, $k_1 = 1$ and $k_2 = 1$, with one change-point at the point $\tau = 90$. The sample sizes are equal to 90 and 160, respectively, before and after the change-point.

Model III is incorrect because it considers that $\alpha = 1$, when in fact it is equal to 2. Models I and II are correct, but Model II incorporates the information that $k_1 = k_2 = 1$, while in Model I this parameter is estimated.

The summaries of the estimates of the posterior distributions of Models I and II were obtained using 100,000 iterations, a burn-in of 30,000 and jumps of 50. The summaries of the restricted Model III are obtained using 80,000 iterations, a burn-in of 20,000 and jumps of 10.

Because these models are very sensitive to the specification of the prior distributions of some parameters, and also because we are now working with more complex models, we performed a sensitivity study for the different specifications of the prior distributions.

Sensitivity study for different specifications of the prior distributions

Vicini et al. (2011) analyzed the sensitivity of the estimates of parameter k_i of Model I using the second set of specifications. They tested various specifications of prior distributions and different values for these hyperparameters because they found difficulties in obtaining convergence of the chains. We do the same here for these parameters and also for the other parameters. Sensitivity analysis was performed and the best result for this model was obtained using the following specifications of the prior distributions of the parameters $\theta_1 \sim \text{Gamma}(0.001, 0.001)$; $\theta_2 \sim \text{Gamma}(0.001, 0.001)$; $\beta_1 \sim \text{Gamma}(1, 1000)$; $\beta_2 \sim \text{Gamma}(1, 100)$; $\alpha_1 \sim \text{Gamma}(2, 1)$; $\alpha_2 \sim \text{Gamma}(1, 1)$ and $\tau \sim \text{Uniform}(80, 100)$. The selection of the hyperparameters is done so that the prior expectation of the parameters is equal to the true parameter values, but the variance, which determines how much information is in the prior distribution, were

taken to be very large .

We only show the graphical results of the chains and the summary of the estimates of the posterior distributions of Model I, since the results for Models II and III are similar.

Figure 6.2 shows the graphs of the chains for the parameters of Model I with one change-point.

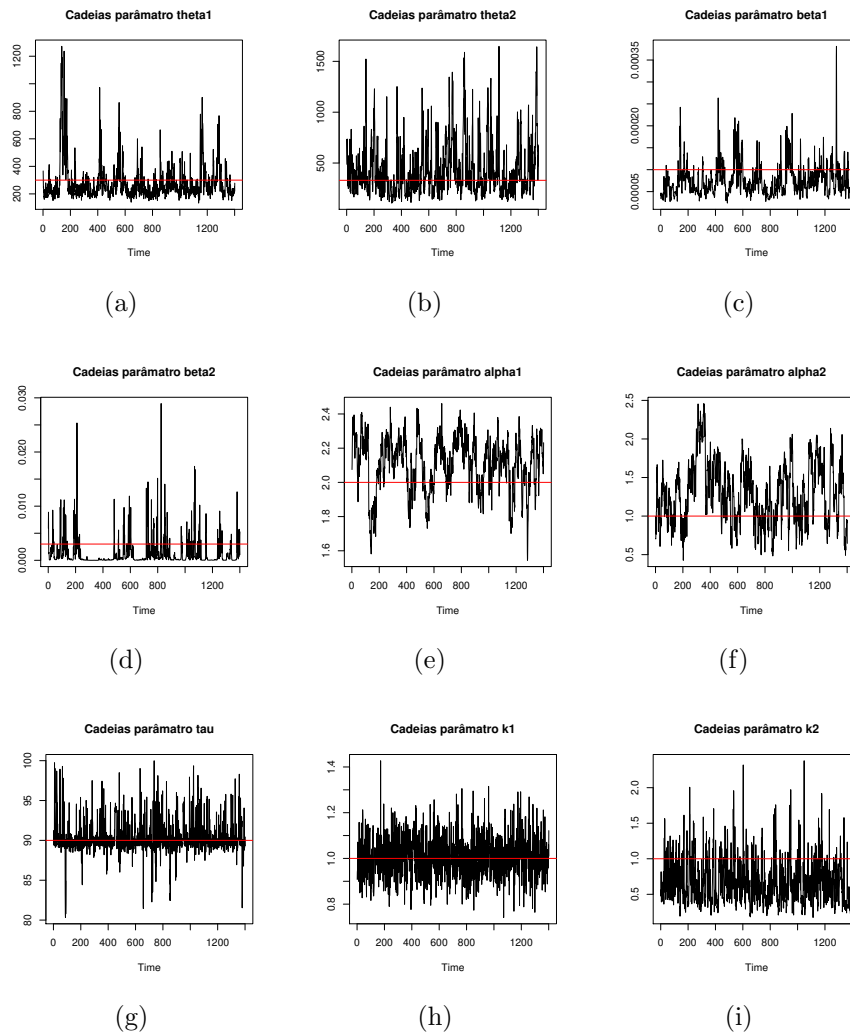


Figura 6.2: Simulation of the chains of model I with one change-point, and the horizontal trace is the true value of each parameter.

In Figure 6.3 we show the posterior distributions, a priori and the true value of each parameter to the model I with one change-point.

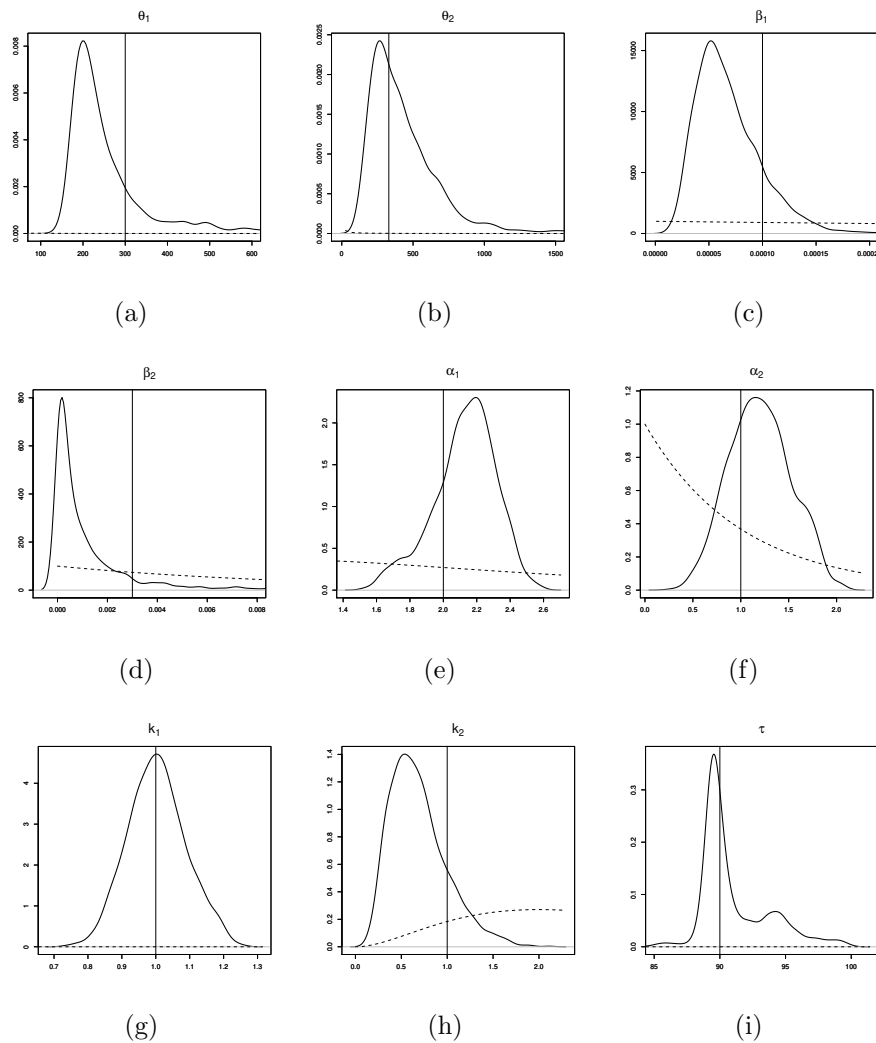


Figure 6.3: Simulation of the posterior distribution of parameters related to Model I, with one change-point. The posterior distributions are shown with a solid line and the prior distributions are shown with a dashed line. The true values of each parameter are represented by vertical lines.

Observing Figures 6.2 and 6.3, for the simulated data set with one change-point, for Model I, we can see that the parameters converge to their true values, as can be seen by the horizontal trace in Figure 6.2 and the vertical line in Figure 6.3. We also note that the specifications of the prior distributions are diffuse, as seen in Figure 6.3, indicating a greater emphasis on the likelihood .

Figure 6.4 shows the adjustment of Models I, II and III in the presence of one

change-point.

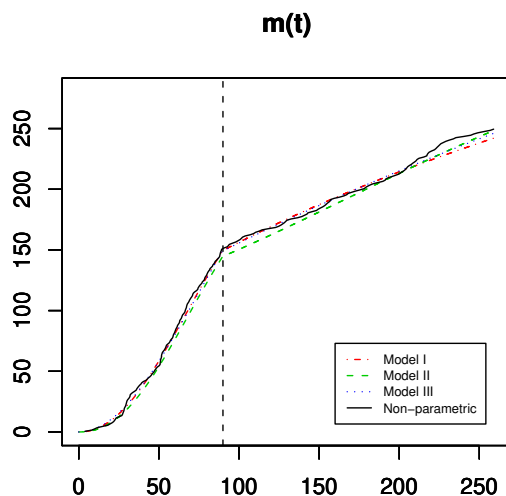


Figure 6.4: Simulated data from the model with one change-point at point $\tau = 90$. Parameters (k, α) are equal to $(1, 2)$ and $(1, 1)$ respectively before and after the change-point. Sample sizes are equal to 90 and 160, respectively before and after the change-point. Graphs of the theoretical function $m(t)$, the empirical estimation and its estimates by models with one change-point: a) Model I, b) Model II c) Model III. The dashed vertical line corresponds to the change-point.

Figure 6.4 presents the graphs of the theoretical function $m(t)$, the empirical estimation and its estimates by Models I, II and III with one change-point. The three models fit the simulated data set with one change-point well. This result can be seen from the curves that are overlap with each other and with the empirical estimate.

Table 6.2 presents summaries of the estimates of the posterior distributions of Models I, II and III with one change-point. The values used were $\theta_1 = 300$, $\theta_2 = 330$, $\beta_1 = 0.0001$, $\beta_2 = 0.003$, $\alpha_1 = 2$, $\alpha_2 = 1$, $\tau = 90$, $k_1 = 1$ and $k_2 = 1$.

We can observe in Table 6.2 that the estimates given by the mean and median of the posterior distributions for the correct Models I and II were satisfactory, because their values are very close to the true values of the parameters; all credibility intervals contain the true value of the parameters; there was no problem of convergence of the MCMC algorithm as observed in the trace plots of the simulated samples; and the estimated

Tabela 6.2: Summary of estimates of the posterior distributions of Models I, II and III with one change-point. Real model: $\theta_1 = 300$, $\theta_2 = 330$, $\beta_1 = 0.0001$, $\beta_2 = 0.003$, $\alpha_1 = 2$, $\alpha_2 = 1$, $\tau = 90$, $k_1 = 1$ and $k_2 = 1$. Sample size of 250. Prior distributions for the parameters θ_1 , θ_2 , β_1 , β_2 , α_1 , α_2 and τ , are presented in Section (6.3). The prior distributions for the parameters k_1 and k_2 in Model I is the exponential distribution with mean parameter equal to 0.99 and truncated at 6, for both parameters.

Model	parameter	mean	median	S.D.	(CI)2.5%	(CI)97.5%
I	θ_1	282.21	240.48	137.03	177.21	540.94
	θ_2	429.51	355.23	258.20	164.26	944.64
	β_1	0.000075	0.000067	0.000035	0.000037	0.000141
	β_2	0.001275	0.000262	0.002503	0.000011	0.006176
	τ	90.42	89.80	2.25	88.49	94.87
	α_1	2.107	2.132	0.159	1.803	2.326
	α_2	1.297	1.281	0.401	0.709	1.993
	k_1	1.004	1.000	0.088	0.867	1.149
	k_2	0.675	0.620	0.315	0.291	1.258
II	θ_1	287.47	223.85	186.94	164.90	629.41
	θ_2	338.33	252.25	238.16	146.10	892.35
	α_1	2.230	2.264	0.260	1.796	2.603
	α_2	1.564	1.613	0.419	0.915	2.132
	β_1	0.0000571	0.0000393	0.0000501	0.0000140	0.0001574
	β_2	0.0006567	0.0001008	0.0016368	0.0000089	0.0029437
	τ	90.92	89.86	2.47	89.05	96.30
III	θ_1	317.20	289.59	105.10	201.35	548.27
	θ_2	687.24	479.50	557.89	283.35	2043.14
	β_1	0.0000936	0.0000899	0.0000345	0.0000406	0.0001538
	β_2	0.0017483	0.0015654	0.0011074	0.0002865	0.003808
	τ	90.21	89.66	1.75	88.85	94.44

curves adjust to the empirical curve very well. The largest value of the Gelman-Rubin's statistics is equal to 1.09, obtained for the parameter θ_2 , also indicating convergence of the MCMC chains. The estimates for Model III, which has the incorrect restriction that $\alpha_1 = 1$, lead to poor estimates. The variances of the posterior distributions of Model II, which incorporates the right information, are not always lower than the variances of the posterior distributions of Model I.

With the sensitivity analysis, we verified that when the proposed model is correct it is not necessary to use informative prior distributions to obtain a posterior distribution with median close to the true values.

6.4.3 Simulated data with two change-points

We simulated 150 observations with parameters given by $\theta_1 = 190$, $\theta_2 = 130$, $\theta_3 = 365$, $\beta_1 = 0.009$, $\beta_2 = 0.007$, $\beta_3 = 0.003$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 1$, $\tau_1 = 50$, $\tau_2 = 100$, $k_1 = 1$, $k_2 = 1$ and $k_3 = 1$, with change-points $\tau_1 = 50$ and $\tau_2 = 100$, with 50 observations before τ_1 , 50 observations between τ_1 and τ_2 and 50 after τ_2 .

Here all three models, Models I, II and III, are correct, but Model II incorporates the information that $k_1 = k_2 = 1$ and Model III incorporates the information that $\alpha_1 = \alpha_2 = \alpha_3 = k_1 = k_2 = k_3 = 1$, while in Model I these parameters are estimated.

The summaries of the estimates of the posterior distributions of Model I are obtained using 100,000 iterations, a burn-in of 30,000 and jumps of 50. For summaries of the estimates of the distributions of Models II and III we used 80,000 iterations, a burn-in of 20,000 and jumps of 50.

We also performed a sensitivity analysis with respect to the specifications of the a priori parameters. Here we present only the best result for the parameter estimates we obtained for these models, which are given using the following specifications of prior distributions for the parameters $\theta_1 \sim \text{Gamma}(0.001, 0.001)$; $\theta_2 \sim \text{Gamma}(0.001, 0.001)$; $\theta_3 \sim \text{Gamma}(0.001, 0.001)$; $\beta_1 \sim \text{Gamma}(1, 100)$; $\beta_2 \sim \text{Gamma}(1, 100)$; $\beta_3 \sim \text{Gamma}(0.01 \times 100, 100)$; $\alpha_1 \sim \text{Gamma}(1, 1)$; $\alpha_2 \sim \text{Gamma}(1, 1)$; $\alpha_3 \sim \text{Gamma}(1, 1)$; $\tau_1 \sim \text{Uniform}(40, 60)$ and $\tau_2 \sim \text{Uniform}(90, 110)$.

We only present the graphical results of Model I, where all the parameters are free.

Figure 6.5 shows the graphs of the chains for the parameters of Model I with two change-points. Figure 6.6 shows the posterior and prior distributions and the true value of each parameter for Model I with two change-points.

The analyses of Figures 6.5 and 6.6 give an indication that the chains are converging to their true values, as can be seen from the horizontal trace in Figure 6.5 and the vertical line in Figure 6.6. Similarly, in Figures 6.2 and 6.3, we also found that the specifications of the prior distributions are diffuse, as seen in Figure 6.6, with likelihood playing a more important role.

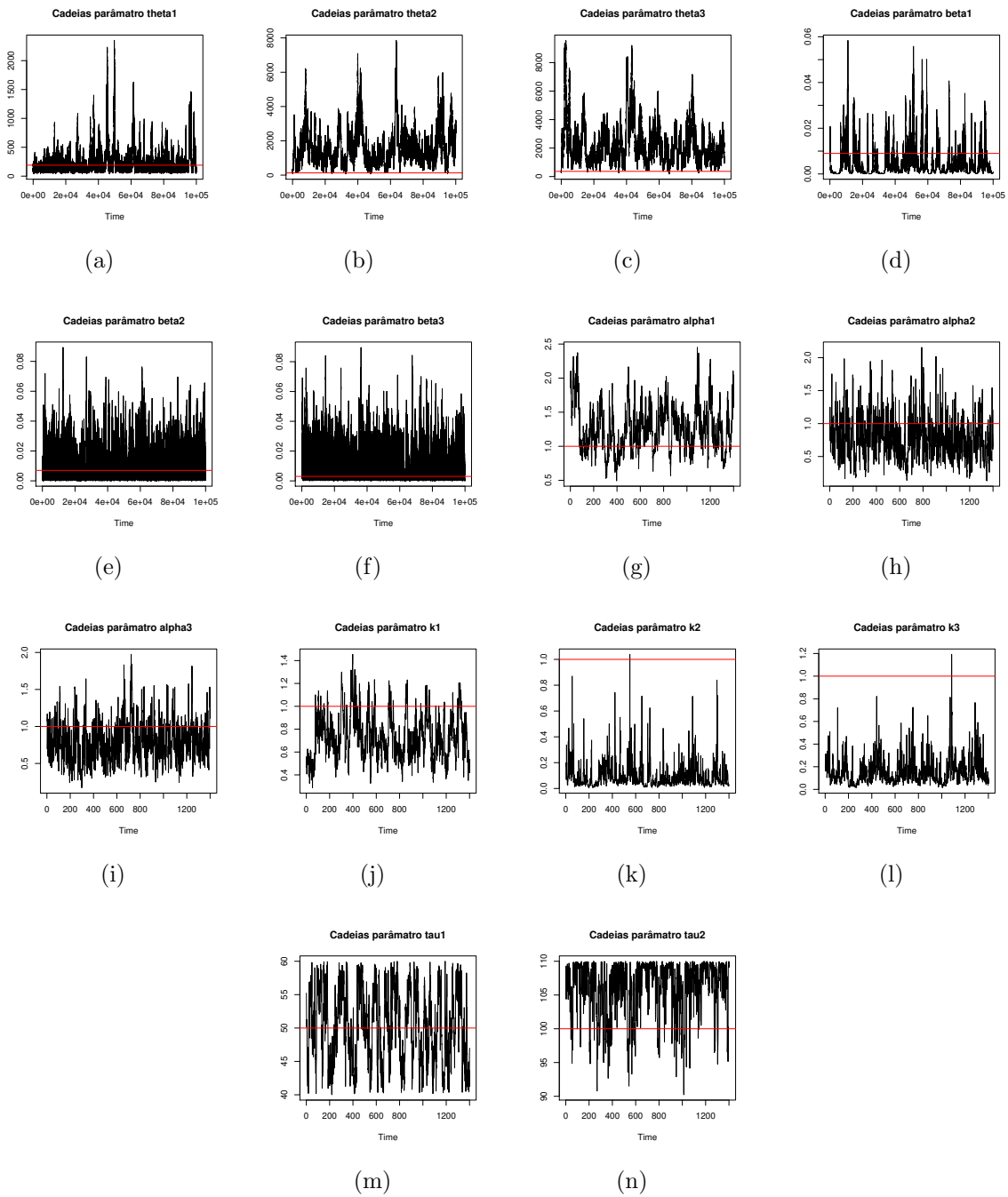


Figura 6.5: Simulation of the chains of Model I with two change-points. The horizontal traces are the true values of each parameter.

Figure 6.7 shows the fit of Models I, II and III, with the presence of two change-points.

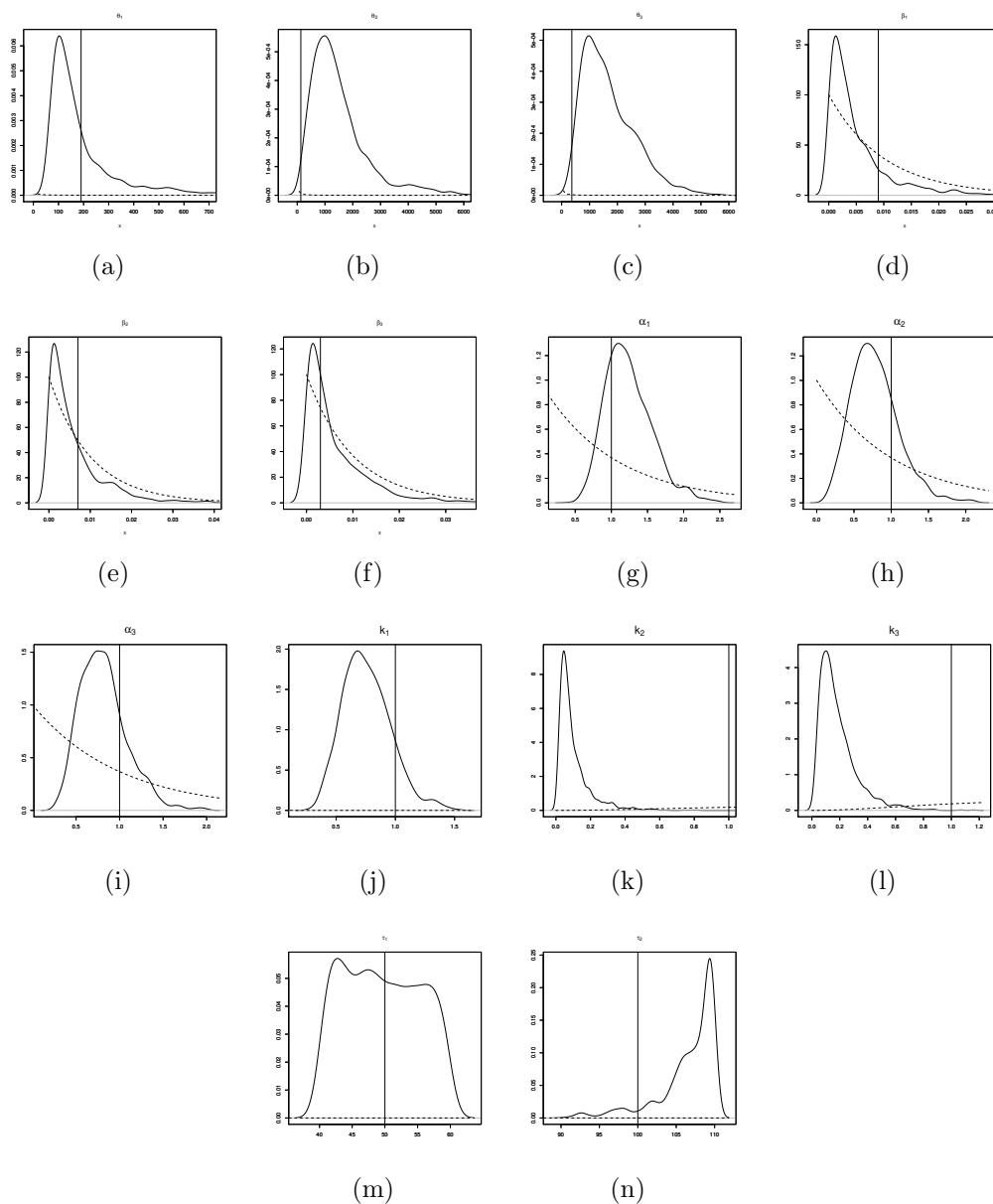


Figure 6.6: Simulation of the posterior distribution of the parameters of Model I with two change-points. The posterior distributions are shown with a solid line and the prior distributions are shown with the dashed line. The true values are indicated by vertical lines.

Figure 6.7 shows the graphs of the function $m(t)$, theoretical and empirical, and its estimation by Models I, II and III. All models fit the simulated data set with two change-points well. This result can be seen from the curves that are overlap with each

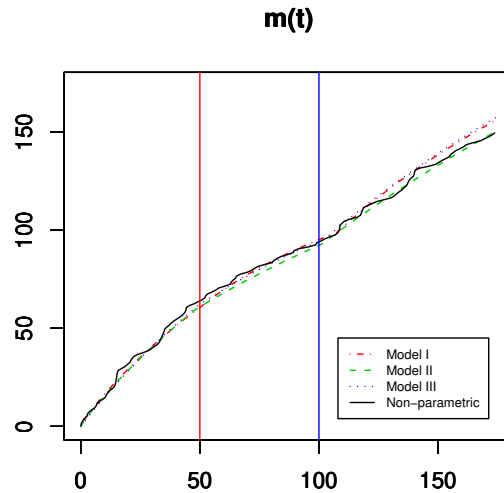


Figure 6.7: Model with two change-points. Sample size of 150. Graphs of the theoretical function $m(t)$, the empirical estimation and its estimates by models with two change-points: a) Model I, b) Model II c) Model III. The vertical lines indicate the change-points.

other.

Table 6.3 presents the estimates of the parameters of Models I, II and III with two change-points. Real model $\theta_1 = 190$, $\theta_2 = 130$, $\theta_3 = 365$, $\beta_1 = 0.009$, $\beta_2 = 0.007$, $\beta_3 = 0.003$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 1$, $\tau_1 = 50$, $\tau_2 = 100$, $k_1 = 1$, $k_2 = 1$ and $k_3 = 1$. Sample size of 150. Prior distributions for the parameters θ_1 , θ_2 , β_1 , β_2 , α_1 , α_2 and τ , are presented in Section 6.3. The prior distributions for the parameters k_1 and k_2 in Model I are the second set (2) proposed, presented in Section 6.3, where we have the exponential prior distribution with mean parameter equal to 0.99 and truncated at 6, for both parameters.

Analyzing Table 6.3 we can observe that the estimates of mean and median of Models I, II and III with two change-points were satisfactory, because their values are very close to the true values of the parameters. All the credibility intervals 95% contain the true value of the parameters. There was no problem in the convergence of the chains of the parameters, as shown by the chains and also by the good fitting of the

curves. The largest value of the Gelman-Rubin's statistics is equal to 1.15, obtained for the parameter θ_2 , also indicating convergence of the MCMC chains. Note that the three models are correct, but that Models II and III incorporate more information, and Model III incorporates more than Model II. We also see, by comparing the results of Model III with the other models, that adding the information that $k_1 = k_2 = k_3$ changes the posterior distribution of the parameter β_1 very little, indicating that information on the parameters k 's is not relevant to the parameter β_1 . Analyzing Tables 6.2 and 6.3 simultaneously, we can see that the amplitude of the credibility intervals of the parameters are larger in Table 6.3, because the sample size is smaller in this table.

Tabela 6.3: Summary of estimates of the posterior distributions for Models I, II and III with two change-points. Real model: $\theta_1 = 190$, $\theta_2 = 130$, $\theta_3 = 365$, $\beta_1 = 0.009$, $\beta_2 = 0.007$, $\beta_3 = 0.003$, $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 1$, $\tau_1 = 50$, $\tau_2 = 100$, $k_1 = 1$, $k_2 = 1$ and $k_3 = 1$. Sample size of 150. Prior distributions for the parameters θ_1 , θ_2 , β_1 , β_2 , α_1 , α_2 and τ , are presented in Section 6.3. The prior distributions for the parameters k_1 and k_2 in Model I are the exponential distribution with the mean parameter equal to 0.99 and truncated at 6, for both parameters.

Model	parameter	mean	median	S.D.	2.5%	97.5%
I	θ_1	244.23	153.65	246.14	70.73	952.61
	θ_2	1587.49	1322.18	1080.72	321.76	4459.74
	θ_3	2121.27	1804.27	1353.79	552.48	5761.59
	β_1	0.00405	0.00221	0.00518	0.00013	0.01768
	β_2	0.00611	0.00368	0.00736	0.00006	0.02716
	β_3	0.00606	0.00366	0.0069	0.00007	0.0255
	τ_1	50.38	50.55	5.50	40.83	59.26
	τ_2	106.00	107.34	4.07	96.44	109.92
	α_1	1.263	1.232	0.335	0.716	2.022
	α_2	0.786	0.741	0.337	0.260	1.570
	α_3	0.797	0.768	0.284	0.336	1.512
	k_1	0.737	0.707	0.198	0.417	1.168
	k_2	0.093	0.063	0.096	0.013	0.339
	k_3	0.151	0.118	0.115	0.030	0.450
II	θ_1	271.77	209.24	205.09	79.88	856.36
	θ_2	456.94	211.64	556.71	78.73	2240.62
	θ_3	650.06	424.86	538.86	192.49	2201.55
	β_1	0.01021	0.00902	0.00578	0.00252	0.02462
	β_2	0.01179	0.00911	0.01005	0.00041	0.03658
	β_3	0.01031	0.00759	0.00963	0.00046	0.0368
	τ_1	49.20	48.45	5.53	40.55	59.20
	τ_2	101.05	101.12	4.73	92.18	108.27
	α_1	0.947	0.932	0.139	0.708	1.255
	α_2	0.867	0.859	0.297	0.320	1.452
α_3	0.867	0.845	0.259	0.431	1.439	

Tabela 6.4: Continuation of Table 6.3.

Model	parameter	mean	median	S.D.	2.5%	97.5%
III	θ_1	221.80	171.96	153.60	85.79	700.27
	θ_2	184.82	143.24	130.11	89.92	598.77
	θ_3	397.48	331.83	226.51	231.97	1179.89
	β_1	0.01001	0.009	0.00567	0.00202	0.02358
	β_2	0.0097	0.0082	0.0069	0.00108	0.02642
	β_3	0.00638	0.00561	0.00412	0.00074	0.01603
	τ_1	50.00	49.85	5.53	40.79	59.13
	τ_2	102.16	102.68	4.55	92.58	108.37

6.5 Application to ozone data

Air pollution is one of the main problems facing large cities. There are many pollutants that cause problems to the population, but what most affects large cities is ozone gas. In some cities, like Mexico City, the authorities have been concerned with the high levels of pollution that it presents. Because of this, the environmental authorities have implemented measures aimed at reducing the level of pollutants. These measures are extremely important, because when pollution levels reach a certain threshold of concentration for a given period of time, individuals exposed to the pollutant can suffer serious health problems (see, for example, Wilson et al., 1980; Loomis et al., 1996; Galizia and Kinney, 1999; Bell et al., 2004, Bell et al., 2005, Bell et al., 2007; Gauderman et al., 2004 and ARB, 2005). In this section, we apply the proposed models to the measurements of ozone in the Metropolitan Area of Mexico City. The Mexican ozone standard is 0.11 parts per million (0.11 ppm) and the threshold used in Mexico City to declare an emergency is 0.22 ppm (see, for example, Achcar et al., 2009). Here we consider a threshold of 0.20 ppm. This value is used because it is between the other two. We apply the models to fit the data corresponding to the maximum daily average measurements of ozone gas, based on data measured in the Northeast region (NE) of Mexico City and a sample of 981 observations, which correspond to times when a certain threshold established for the air quality standard is violated during the period of time T which we considered. These data are collected from www.sma.df.gob.mx/simat/, which account for about 18 years of observations

(1 January 1990 to 31 December 2008).

The summaries of the estimates of the posterior distributions of Model I with no change-points were obtained using 700,000 iterations, a burn-in of 15,000 and jumps of 100. The specification of the prior distributions of the parameters for this model were $\theta \sim \text{Gamma}(0.001, 0.001)$; $\beta \sim \text{Gamma}(10, 100)$; $\alpha \sim \text{Gamma}(0.01, 0.01)$. The prior distribution for the parameter k is the exponential distribution with the mean parameter equal to 0.95 and truncated at 3.

For Model I with one change-point, the posterior estimations are obtained using 400,000 iterations, a burn-in of 200,000 and jumps of 100. The specifications of the prior distributions of the parameters for this model were $\theta_1 \sim \text{Gamma}(0.001, 0.001)$; $\theta_2 \sim \text{Gamma}(0.001, 0.001)$; $\beta_1 \sim \text{Gamma}(0.001, 0.001)$; $\beta_2 \sim \text{Gamma}(0.001, 0.001)$; $\alpha_1 \sim \text{Gamma}(1, 1)$; $\alpha_2 \sim \text{Gamma}(1, 1)$ and $\tau_1 \sim \text{Uniform}(2900, 2950)$. The prior distributions for the parameters k_1 and k_2 are the exponential distribution with mean parameter 0.99 and truncated at 6 for both parameters.

For Model I with two change-points the posterior estimations were obtained using 400,000 iterations, a burn-in of 300,000 and jumps of 200. The specifications of the prior distributions of the parameters for this model were $\theta_1 \sim \text{Gamma}(0.001, 0.001)$; $\theta_2 \sim \text{Gamma}(0.001, 0.001)$; $\theta_3 \sim \text{Gamma}(0.001, 0.001)$; $\beta_1 \sim \text{Gamma}(6, 1000)$; $\beta_2 \sim \text{Gamma}(1, 1000)$; $\beta_3 \sim \text{Gamma}(1, 100)$; $\alpha_1 \sim \text{Gamma}(1, 1)$; $\alpha_2 \sim \text{Gamma}(1, 1)$; $\alpha_3 \sim \text{Gamma}(1, 1)$; $\tau_1 \sim \text{Uniform}(2900, 2950)$ and $\tau_2 \sim \text{Uniform}(4340, 4360)$. The prior distributions for the parameters k_1 , k_2 and k_3 are the exponential distribution with mean parameter 0.99 and truncated at 6 for all three parameters.

Figure 6.8 presents the mean function $m(t)$, theoretical and empirical, and its estimates by the models with no change-points and with one and two change-points. We can observe that the graphs of the function $m(t)$ estimated by the models without any change-points and with one and two change-points fit the data set very well. This result can be observed in the adjusted curves, which are overlap the empirical curve.

Table 6.5 presents a summary of estimates of the posterior distributions for the models without any change-points and with one and two change-points for Model I.

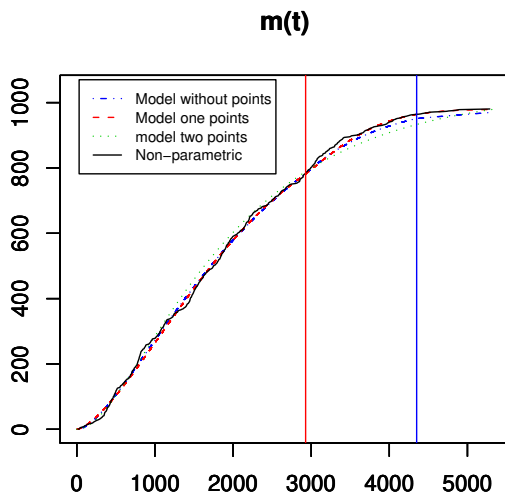


Figura 6.8: Graphs of the mean value of the function $m(t)$, theoretical and empirical, and its estimates by the models without any change-points and with one and two change-points. Ozone pollution from NE region of Mexico City. The vertical lines show the estimated change-points.

Prior distributions for the parameters θ_i , β_i , α_i and τ_j , are presented in Section 6.3. The prior distributions for the parameters k_i , are those of the second set (2) proposed, in which we have the prior exponential distribution with mean parameter 0.95 truncated at 3 for the model without change-points and with mean parameter 0.99 truncated at 6 for the models with one and two change-points.

Table 6.5 presents the posterior estimates of the parameters for the models without any change-points and with one and two change-points for Model I.

We also use the Deviance Information Criterion (DIC) proposed by Spiegelhalter et al. (2002) to compare the models. This criterion is widely used in Bayesian model selection when samples of posterior distributions of parameters are obtained by MCMC simulation. Table 6.6 shows the values of DIC for the models with no change-points and with one and two change-points. The DIC selects the more complex models with two change-points.

Tabela 6.5: Summaries of estimates of the posterior distributions for the models without any change-points and with one and two change-points for Model I. Prior distributions for the parameters θ_i , β_i , α_i and τ_j , are presented in Section 6.3. The prior distributions for the parameters k_i are the exponential distribution with mean parameter equal to 0.95 and truncated at 3 for the model without change-points and mean equal to 0.99 and truncated at 6 for the models with one and two change-points.

Model	parameter	mean	median	S.D.	2.5%	97.5%
without change-points	θ	1035.04	1034.45	35.20	979.42	1093.69
	β	0.0105081	0.0104119	0.0022471	0.0068893	0.0144018
	α	0.744	0.743	0.028	0.700	0.793
	k	2.898	2.925	0.093	2.710	2.995
one change-point	θ_1	994.25	986.88	62.87	903.48	1105.67
	θ_2	348.85	344.48	44.87	282.13	426.69
	β_1	0.0000100	0.0000097	0.0000025	0.0000067	0.0000146
	β_2	0.0000052	0.0000028	0.0000069	0.0000003	0.0000175
	τ	2923.45	2922.35	13.51	2902.91	2946.19
	α_1	1.501	1.502	0.038	1.437	1.561
	α_2	1.810	1.809	0.139	1.591	2.086
	k_1	1.000	0.999	0.045	0.929	1.075
k_2	5.982	5.987	0.019	5.945	5.999	
two change-points	θ_1	1175.47	1166.73	107.61	991.77	1403.03
	θ_2	3416.58	3215.40	1380.97	1331.79	6816.79
	θ_3	2354.08	1951.01	1448.58	601.19	5735.09
	β_1	0.00504	0.00471	0.00221	0.00167	0.0105
	β_2	0.00048	0.00038	0.00042	0.00007	0.00154
	β_3	0.0059	0.00381	0.00613	0.00015	0.023
	τ_1	2924.12	2923.43	13.40	2902.10	2946.76
	τ_2	4350.87	4350.88	4.99	4341.21	4359.29
	α_1	0.796	0.793	0.051	0.709	0.914
	α_2	1.110	1.104	0.111	0.890	1.314
	α_3	0.527	0.518	0.152	0.273	0.853
	k_1	2.298	2.290	0.213	1.917	2.730
	k_2	1.044	0.936	0.571	0.320	2.600
k_3	0.186	0.139	0.168	0.026	0.597	

Tabela 6.6: Deviance Information Criterion for models without change-points and with one and two change-points, using a data set of ozone pollution from the NE region of Mexico City.

Models	no change-points	one change-point	two change-points
DIC	4863.141	-2217.76	-2413.32

6.6 Final Remarks

We consider the problem of modeling data on pollution in Mexico City. Because the intensity function may increase or decrease, depending on intervention measures for the air pollution, we use an NHPP process with change-points to model the data.

The parameter estimation is carried out using a formulation based on Bayesian MCMC methods. Sensitivity tests to prior distributions were carried out and showed that the posterior distributions are robust.

We also provide a sensitivity analysis using different parameters for the prior distributions, and study the convergence of the MCMC chains. To illustrate the proposed method we considered simulated data and a data set for air pollution from the NE region of Mexico City.

The analysis was performed in two stages. Initially we worked with models without the presence of change-points; these models are of fundamental importance, because the information obtained from them serves to support more complex models which use change-points. The use of the information obtained in previous stages of the Bayesian analysis is important to achieve convergence of the MCMC chains.

The application of the generalized gamma model using Bayesian inference for NHPP with change-points proved to be an excellent tool to analyze the air pollution caused by ozone gas. It was observed that when applying more complex models, with the inclusion of change-points, the improvement in the fit tended to compensate for the increase in the number of parameters. This is also confirmed by the graphs of the MCMC chains of the parameters, the fitting of the curves and the DIC.

It was observed that the change-points occur after a period during which environmental action was taken by the government to reduce pollution levels. For example, from 1990 to 1998 there was a reduction in the use of highly polluting vehicles and, around this period, the model detected a change-point. This was also observed in Achcar et al. (2009). It is noteworthy that, from 1999 to 2002, a series of measures was taken by the environmental authorities in Mexico, with the goal of reducing ozone

levels; these measures included manufacturers being encouraged to produce cars with modern clean technology. 2001 was a key date here. Matching this, around 2002 another turning point was detected.

Referências Bibliográficas

- Achcar, J. A.; Dey, K.D. e Niverthi, M. – A Bayesian approach using nonhomogeneous poisson process for software reliability models, In: A.S.Basu; S.K.Basu; S.Mukhopadhyay, (Org.). *Frontiers in Reliability*. ed.1 River Edge, NJ: World Scientific, v. 4, p. 1–18, 1998.
- Achcar, J. A.; Fernández-Bremauntz, A. A.; Rodrigues, E.R. e Tzintzun, G. – Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model, *Environmetrics*, v. 19, p. 469–485, 2008.
- Achcar, J. A.; Rodrigues, E.R.; Paulino, C. D. and Soares, P. – Non-homogeneous Poisson models with a change-point: an application to ozone peaks in Mexico City, *Environmental and Ecological Statistics*, v. 17, p. 521–541, 2009.
- Achcar, J. A.; Rodrigues, and Tzintzunc, G. – Using non-homogeneous Poisson models with multiple change-points to estimate the number of ozone exceedances in Mexico City, *Environmetrics*, v. 22, p. 1–12, 2009.
- Achcar, J. A.; Rodrigues, O.G. e Rodrigues, E.R. – The behaviour of a Metropolis-Hastings algorithm under different prior distributions: an application to ozone measurements in Mexico City, 3rd WSEAS International Conference on Urban Planning and Transportation (UPT'10), Corfu Island, Greece, p. 160–165, 2010.
- Achcar, J. A.; Rodrigues, E.R. e Tzintzun, G. – Modelling inter-occurrence times between ozone peaks in Mexico City in the presence of multiple change-points, *Brazilian Journal of Probability and Statistics*, v. 25, n. 2, p. 183–204, 2011.

Air Resource Board (ARB) – Review of air quality standard for ozone in California, Environmental Protection Agency, Staff Report, California, USA, 2005.

Álvarez, L.J.; Fernández-Bremauntz, A.A.; Rodrigues, E.R. e Tzintzun, G. – Maximum a posteriori estimation of the daily ozone peaks in Mexico City, *Journal of Agricultural, Biological and Environmental Statistics*, v. 10, n. 3, p. 276–290, 2005.

Austin, J. e Tran, H. – A characterization of the weekday-weekend behavior of ambient ozone concentrations in California, *Air pollution VI*, WIT Press, Ashurst Lodge, p. 645–661, 1999.

Bell, M.L.; Mcdermontt, A.; Zeger, S.L.; Samet, J.M. e Dominici, F. – Ozone and short-term mortality in 95 US urban communities, *Journal of the American Medical Society*, v. 292, p. 2372–2378, 2004.

Bell, M.L.; Peng, R. e Dominici, F. – The exposure-response curve for ozone and risk of mortality and the adequacy of current ozone regulations, *Environmental Health Perspectives*, v. 114, p. 532–536, 2005.

Bell, M.L.; Goldberg, R.; Rogrefe, C.; Kinney, P.L.; Knowlton, K.; Lynn, B.; Rosenthal, J.; Rosenzweweig, C. e Patz, J.A. – Climate change, ambient ozone, and health in 50 US cities, *Climate Change*, v. 82, p. 61–76, 2007.

Braga, A.L.F.; Zanobetti, A. e Schwartz, J. – The effect of weather on respiratory and cardiovascular deaths in 12 U.S. Cities, *Enviromental Health Perspectives*, v. 110, p. 859–863, 2002.

Casella, G. e Berger, R. L. – *Statistical Inference*, Duxbury Press, Belmont, CA, Second Edition, 2001.

Cox, D.R. e Lewis, P.A. – *Statistical Analysis of Series of Events*, Methuen, 1966.

- Flaum, J.B.; Rao, S.T. e Zurbenko, I.G. – Moderating influence of meteorological conditions on ambient ozone concentrations, *Journal of the Air and Waste Management Association*, v. 46, p. 33–46, 1996.
- Farhat, S.C.L.; Paulo, R.L.P.; Shimoda, L.M.; Conceição, G.M.S.; Lin, C.A.; Braga, A.L.F.; Warth, M.P.N. e Saldiva, P.H.N. – Effect of air pollution on pediatric respiratory emergency room visits and hospital admissions, *Brazilian Journal of Medical and Biological Research*, v. 38, p. 227–235, 2005.
- Galizia, A. e Kinney, P.L. – Long-term residence in areas of high ozone: associations with respiratory health in a nationwide sample of nonsmoking young adults, *Environmental Health Perspectives*, v. 107, p. 675–679, 1999.
- Gauderman, W.J.; Avol, E.; Gililand, F.; Vora, H.; Thomas, D.; Berhane, K.; McConnel, R.; Kuenzli, N.; Lurmmann, F.; Rappaport, E.; Margolis, H.; Bates, D.; Peter, J. – The effects of air pollution on lung development from 10 to 18 years of age, *New England Journal of Medicine*, p. 1057–1067, 2004.
- Iman, A. Inference and monitoring convergence. *Markov Chain Monte Carlo in Practice* (eds. W. R. Gilks, S. Richardson e D. J. Spiegelhalter), London: Chapman and Hall, p. 131–143, 1996.
- Iman, A. and Rubin, D. R. A single series from the gibbs sampler provides a false sense of security. *Bayesian Statistics 4* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith), Oxford: Oxford University Press, p. 625–631, 1992.
- Goel, A.L. e Okumoto, K. – An Analysis Of Recurrent Software Errors In A Real-Time Control System, In Proceedings of ACM Conference, Washington, D.C., USA, p. 496–501, 1979.
- Goel, A.L. – A guidebook for software reliability assessment, Technical Report, RADC-TR-83-176, 1983.

- Gouveia, N.; Freitas, C.U.; Martins, L.C. e Marcilio, I.O. – Hospitalizações por causas respiratórias associadas à contaminação atmosférica no município de São Paulo, Brasil, *Cadernos Saúde Pública*, Rio de Janeiro, v. 22, p. 2669–2677, 2006.
- Guardani, R.; Nascimento, C.A.O; Guardani, M.L.G.; Martins, M.H.R.B. e Romano, J. – Study of atmospheric ozone formation by means of a neural network based model, *Journal of the Air and Waste Management Association*, v. 49, n. 3, p. 316–323, 1999.
- Hastings, W. K. – Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, v. 57, p. 97–109, 1970.
- Horowitz, J. – Extreme values from a nonstationary stochastic process: an application to air quality analysis, *Technometric*, v. 22, p. 469–482, 1980.
- Kou, L, e Yang, T.Y. – Bayesian computation for nonhomogeneous Poisson process in software reliability, *Journal of the American Statistical Association*, v. 91, p. 763–773, 1996.
- Lanfredi, M. e Macchiato, M. – Searching for low dimensionality in air pollution time series, EPL, *Europhysics Letters*, v. 40, p. 589–594, 1997.
- Larsen, L.C.; Bradley, R.A. e Honcoop, G.L. – *A new method of characterizing the variability of air quality-related indicators*, In: Air and waste management association's international specialty conference of tropospheric ozone and the environment, Los Angeles, EUA, 1990.
- Leadbette, M.R. – On a basis for peak over threshold modeling, *Statistics and Probability Letters*, v. 12, p. 357–362, 1991.
- Leemis, L. M. – Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process, *Management Science*, v. 37, p. 886–900, 1991.

- Loomis, D.P.; Borja-Arbutto, V.H.; Bangdiwalla, S.I. e Shy, C.M. – *Ozone exposure and daily mortality in Mexico City: a time series analysis*, Health Effects Institute Research Report, v. 75, p. 1–46, 1996.
- Martins, L.C.; Latorre, M.R.D.O.; Cardoso, M.R.A.; Gonçalves, F.L.T.; Saldiva, P.H.N. e Braga, A.L.F. – Poluição atmosférica e atendimentos por pneumonia e gripe em São Paulo, Brasil, *Revista de Saúde Pública*, v. 36, p. 88–94, 2002.
- Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. – Equations of state calculations by fast computing machines, *Journal of Chemical Physics*, v. 21, p. 1087–1092, 1953.
- Ohba, M.; Yamada, S.; Takeda, K. e Osaki, S. – *S-shaped software reliability growth curve: how good is it?*, COMPSAC’82, p. 38–44, 1982.
- Pan, J-N. e Chen, S-T – Monitoring long-memory air quality data using ARFIMA model, *Environmetrics*, v. 19, p. 209–219, 2008.
- Piegorsch, W.W.; Smith, E.P.; Edwards, D. e Smith, L. – Statistical advances in environmental science, *Statistical Science*, v. 13, p. 186–208, 1998.
- Roberts, E.M. – Review of statistics of extreme values with applications to air quality data, Part I, *Journal of the Air Pollution Control Association*, Pittsburg, v. 29, p. 632–637, 1979a.
- Roberts, E.M. – Review of statistics extreme values with applications to air quality data, Part II. Applications, *Journal of the Air Pollution Control Association*, Pittsburg, v. 29, p. 733–740, 1979b.
- Smith, R.L. – Extreme values analysis of environmental time series: an application to trend detection in ground-level ozone, *Statistical Science*, v. 4, p. 367–393, 1989.
- Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P. e Van der Linde, A. – Bayesian measures of

- model complexity and fit (with discussion), *Journal of the Royal Statistical Society*, B(64), p. 583–639, 2002.
- Vicini, L.; Hotta, L. K. e Achcar, J. A. – *Non-homogeneous Poisson processes applied to count data: a Bayesian approach considering different prior distributions*, submetido em 2011.
- Wilson, R., Colome, S.D., Spengler, J.D. e Wilson, D.G. – *Health Effects in Fossil Fuel Burning: Assessment and Mitigation*, Cambridge, USA: Ballinger, 1980.
- Yamada, S.; Ohba, M. e Osaki, S. – S-Shaped reliability growth modeling for software error detection, *IEEE Transactions on Reliability*, v. 32, p. 475–478, 1983.
- Yang, T.Y. – *Computacional approaches to Bayesian inference for software reliability*, Ph.D. Thesis, University of Connecticut, Storrs, 1994.
- Zolghadri, A. e Henry, D. – Minmax statistical models for air pollution time series. Application to ozone time series data measured in Bordeaux, *Environmental Monitoring and Assessment*, v. 98, p. 275–294, 2004.

Capítulo 7

Estimating rates of nonhomogeneous Poisson Process with multiples change-points using splines

Guilherme Ludwig, Lorena Vicini and Luiz K. Hotta

Abstract. In this paper we introduce the use of functional data analysis to estimate the rate function for a nonhomogeneous Poisson Process. We assume the rate function is a continuous function but with a finite number of discontinuities on its first derivative, at so-called “change points”. The rate function, and its change points, will be estimated with smoothing splines, with a penalization based on the current change point candidates. The developed methods will be tested by means of simulation and employed to study the ozone pollution data from Mexico City, consisting of the number of times the air quality standard, based on ozone concentration levels, is exceeded in a given period of time.

7.1 Introduction

Functional data analysis provides tools for prediction and for estimation of the moments, curvature, derivatives and other characteristics of the underlying

distribution that the functional data has. Spline based methods assume only that the data are a realization of a stochastic process in a smooth functional space. A thorough reference to splines is the book by de Boor (2001), from a numerical analysis viewpoint. Wahba (1990), Ferraty and Vieu (2006) and Ramsay and Silverman (2005) are focused on the statistical use of spline-based methods, such as smoothing splines, functional nonparametric regression and principal components analysis.

When an abrupt change of regimen in the rate function $\lambda(t)$ happens at the point c , c will be called a *change point*. Nonparametric analysis of nonhomogeneous Poisson processes' rate function $\lambda(t)$ was first proposed by Leemis (1991), and a spline-based methodology was shown by Ramsay and Silverman (2005). Change point estimation, with regard to the parametric scenario, can be reviewed in the work by Achcar, Rodrigues, Paulino and Soares (2009) and references therein. Our proposal is that one may understand a change point in the spline based model as a discontinuity in the first derivative of $\lambda(t)$, thus unifying these two approaches to the problem.

Our work aims to simultaneously estimate the change points of nonhomogeneous Poisson processes with smoothing splines and the rate function itself. This paper is organized as follows Section 7.2 introduces the methodologies employed in the estimation, namely, spline functions (and their nonnegative, exponentiated version), smoothing splines and their relationship with derivatives of the estimates, and how this relationship can be extended to more general differential operators. In the Section 7.3 a simulation study was conducted to assess the performance of the proposed methods. An application to Mexico City's pollution data is performed in Section 7.4. Finally, Section 7.5 discusses the results and the open problems related to these proposals.

7.2 Nonhomogeneous Poisson Process Model for Functional Data

7.2.1 Spline functions

A functional data $y(t)$ is defined as the realization of a random function $Y(t)$, such that, for any collection of n points t_1, t_2, \dots, t_n , it is characterized by

$$\mathbb{P}(Y(t_1) \leq y_1, \dots, Y(t_n) \leq y_n) = F_{t_1, t_2, \dots, t_n}(y_1, \dots, y_n).$$

Such a definition is due to Yaglom (1987). If we assume that

$$\mathbb{E}(Y(t)) = \alpha(t),$$

where $\alpha(t)$ is a smooth function. Then, by an estimate of the functional data, we mean the expansion of $\alpha(t)$ into a linear combination of K known basis functions $B_k(t)$, and corresponding coefficients β_k , which are the parameters to be estimated. That is,

$$\alpha(t) = \sum_{k=1}^K \beta_k B_k(t),$$

or equivalently,

$$Y(t) = \sum_{k=1}^K \beta_k B_k(t) + \varepsilon(t),$$

where we assume $\varepsilon(t)$ to be some error with zero mean and finite variance. In this article we will use the *B-splines* as basis functions (Ramsay and Silverman 2005). Given a set τ_1, \dots, τ_L of points in a closed interval, called *knots*, an order m *B-spline* $B_k(t)$, for $k = 1, \dots, K$, is an order m polynomial on each interval $[\tau_\ell, \tau_{\ell+1}]$, which is $m - 1$ differentiable on the knots other than τ_1 and τ_L , and which can be obtained via a recursive relation (see de Boor 2001, Ramsay and Silverman 2005). They form a basis for the spline functions on the interval $[\tau_1, \tau_L]$.

The estimate for the mean rate function $\lambda(t)$ through basis function expansion for the nonhomogeneous Poisson processes was shown in Ramsay and Silverman (2005, p. 122). Such estimation is a special case of *constrained estimation*, in the sense that the final estimates must be nonnegative functions. To achieve this result, the bases are simply exponentiated before being introduced into the likelihood function of the Poisson process. In other words, we take a $\lambda(t)$ such that

$$\lambda(t) = \exp \left\{ \sum_{k=1}^K \beta_k B_k(t) \right\} = \exp\{\boldsymbol{\beta}'\mathbf{B}(t)\}, \quad (7.1)$$

where $B_k(t)$ are the B-spline basis, and β_k the coefficients to be estimated (we will use the notation $\boldsymbol{\beta}'\mathbf{B}(t)$ as a shorthand for their linear combination henceforth). The log-likelihood function of the Poisson process with rate given by (7.1) will have the form

$$\mathcal{L}(\boldsymbol{\beta} \mid t_1, t_2, \dots, t_n, T) = \sum_{i=1}^n \log(\lambda(t_i)) - \int_0^T \lambda(s) ds = \sum_{i=1}^n \boldsymbol{\beta}'\mathbf{B}(t_i) - \int_0^T \exp\{\boldsymbol{\beta}'\mathbf{B}(t)\} ds.$$

The problem is now reduced to finding $\boldsymbol{\beta}$ such that \mathcal{L} is maximized, which can be obtained numerically.

7.2.2 Spline Smoothing and Differential Operators

The spline estimation methods possess interesting tools such as smoothing and the possibility of fitting differential equation models to the data. Roughness smoothing is discussed in Ramsay and Silverman (2005, p. 81) and Wahba (1990), among others. It consists of penalizing estimates with a large second derivative norm, and the estimating procedure thus yields smoother estimates. That is, if we define

$$\text{PEN}_2(\lambda) = \int_{-\infty}^{\infty} \left[\frac{\partial^2}{\partial s^2} \lambda(s) \right]^2 ds, \quad (7.2)$$

then we can add the penalization $\text{PEN}_2(\lambda)$, multiplied by a smoothing parameter ξ to the fitting criterion data. This forces the estimates of β_k to produce smooth estimates

of $\lambda(t)$. While there are many ways to find the tuning parameter ξ , it is not easy to minimize a GCV-like criterion, as in the spline regression case (detailed in Wahba 1990, Ramsay and Silverman 2005). Instead, we choose ξ in an *ad hoc* manner, comparing the nonpenalized likelihood against its penalized version. Such a naïve procedure follows the rules of thumb suggested by Ramsay and Silverman (2005, p. 95).

7.2.3 Detecting Change Points by Differential Equations

Differential equation fitting is a natural extension of the smoothing method, presented in Ramsay and Silverman (2005) and references therein. Consider a linear differential equation given by

$$y^{(p)}(t) = -\gamma_0 y(t) - \gamma_1 y'(t) - \dots - \gamma_{(p-1)} y^{(p-1)}(t) + u(t), \quad (7.3)$$

which can be rewritten as

$$L_\gamma y(t) = \gamma_0 y(t) + \gamma_1 y'(t) + \dots + \gamma_{(p-1)} y^{(p-1)}(t) + y^{(p)}(t) - u(t) = 0 \quad (7.4)$$

We define

$$\text{PEN}_L(\lambda) = \int_{-\infty}^{\infty} [L\lambda(s)]^2 ds,$$

where L is a differential operator which annihilates the structure of a differential equation, or in other words, $L\lambda^* = 0$ if, and only if, λ^* is a solution of the differential equation (7.3). This penalization brings the estimate $\hat{\lambda}$ closer to being one of the solutions λ^* . Assumptions regarding the adequacy of spline-based methods are discussed in Lyche and Schumaker (1994). The existence of solutions, for the general case, are discussed in Ramsay, Hooker, Campbell and Cao (2007), and our linear case is discussed in Ramsay (2000). Further theory for differential equation based on functional data fitting can be found in Appendix .1, in the paper by Ramsay et al. (2007), or in Chapter 19 of Ramsay and Silverman (2005).

The differential equation whose solution is given by a straight line of inclination γ_1

up to the change point c , and afterwards having inclination γ_2 , is given by

$$\lambda'(t) = \gamma_1(1 - \Delta(t - c)) + \gamma_2\Delta(t - c),$$

where $\Delta(t)$ is the indicator function for the continuous set $\{t > 0\}$, and $\lambda'(t) = \frac{d}{dt}\lambda(t)$.

This equation can be rewritten as

$$L_\gamma\lambda(t) = \lambda'(t) - \gamma_2\Delta(t - c) - \gamma_1(1 - \Delta(t - c)), \quad (7.5)$$

and the explicit solution to $L_\gamma\lambda(t) = 0$, subject to the initial condition $y(c) = \gamma_0$, is

$$\lambda(t) = \gamma_0 + \gamma_1(c - x)(1 - \Delta(x - c)) + \gamma_2(x - c)\Delta(x - c).$$

Our proposal is that one may combine those techniques to estimate the Poisson process rate function λ in the presence of change points. The joint likelihood fitting and differential equation fidelity criterion is given by

$$\mathcal{L}_{L_\gamma}(\boldsymbol{\beta}, \xi \mid t_1, t_2, \dots, t_n, T) = \sum_{i=1}^n \boldsymbol{\beta}'\mathbf{B}(t_i) - \int_0^T \exp\{\boldsymbol{\beta}'\mathbf{B}(s)\} ds - \xi \int_0^T (L_\gamma\lambda(t))^2 dt \quad (7.6)$$

The estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, for a fixed ξ , are obtained by

$$\left(\hat{\boldsymbol{\beta}}_\xi, \hat{\boldsymbol{\gamma}}_\xi \right) = \arg \max_{\boldsymbol{\beta}, \boldsymbol{\gamma}} \mathcal{L}_{L_\gamma}(\boldsymbol{\beta} \mid t_1, t_2, \dots, t_n, T),$$

and the $\lambda(t)$ rate estimate is thus

$$\hat{\lambda}(t) = \exp \left\{ \sum_{k=1}^K \hat{\beta}_k B_k(t) \right\}.$$

The first derivative of the nonhomogeneous Poisson process rate function, given by

exponentiated splines, is given by

$$\lambda'(t) = \left(\sum_{k=1}^K \beta_k \mathbf{B}'_k(t) \right) \exp \left\{ \sum_{k=1}^K \beta_k \mathbf{B}_k(t) \right\}, \quad (7.7)$$

and therefore the operator (7.5) is explicitly given by

$$\|L_\gamma \lambda(t)\| = \int_{-\infty}^{\infty} \left(\left(\sum_{k=1}^K \beta_k \mathbf{B}'_k(t) \right) \exp \left\{ \sum_{k=1}^K \beta_k \mathbf{B}_k(t) \right\} - \gamma_2 \Delta(t-c) - \gamma_1 (1 - \Delta(t-c)) \right)^2 dt.$$

Such penalization is the cornerstone of our work: it can incorporate information about a change point c – if it is known – yielding estimates that fit the data better (since such first derivative discontinuity would be a small violation of the smoothness hypothesis of spline-based procedures) and, furthermore, it can be used to find one or more *unknown* change points by examining a data fitting criterion like the AIC (we refer to Dias and Garcia, 2007, and references therein for a justification of AIC based methods when using splines, although their paper puts the smoothness problem in the context of number of basis K selection, instead of the PEN norm).

7.3 Simulated Example

First we show how ordinary splines and ordinary smoothing splines work in a simulation scenario. Figure 7.1 shows which rate function is sampled from: it is the solid, black line with a nondifferentiable point in $t = 3500$ – its change point. The blue line represents the ordinary exponentiated cubic spline estimate, with 8 internal knots, equally spaced in the interval $[0, 10000]$. Applying the roughness smoothing method to simulated data, with parameter $\xi = 10^{10}$ (which gives a penalty of about 2×10^1 against a likelihood of 6×10^3) and the same eight internal knots, we obtain the red curve.

The PEN_L term for the simulated data, using a single change point, is roughly of order 10^{-4} when the likelihood is examined for a initial value of $\beta^{(0)}$ given by the

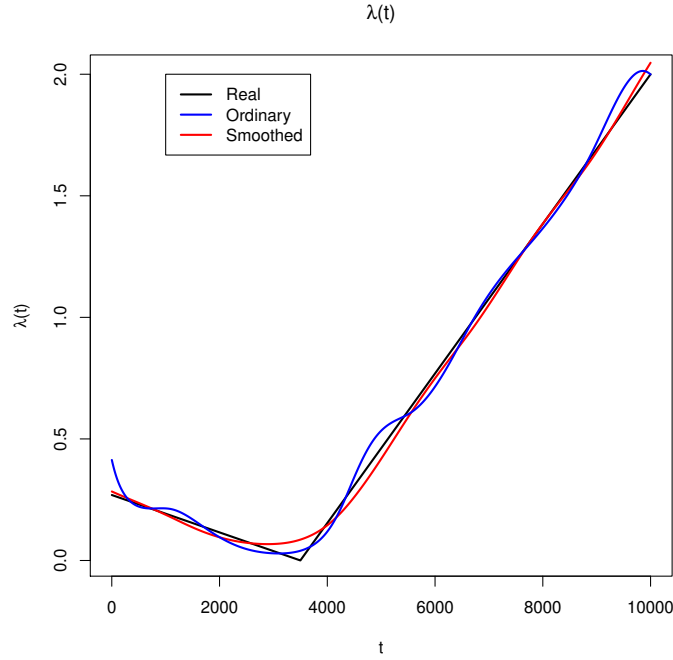


Figure 7.1: Maximum likelihood estimate (colored blue) and its roughness penalized version for the simulated data (in red)

ordinary spline fit. Since the likelihood is of order 10^3 , we picked $\xi = 10^6$ and $\xi = 10^7$ for tuning parameter candidates. The parameters γ and the coefficients β are obtained iteratively. First we obtain a maximum likelihood estimate $\hat{\beta}^{(0)}$, using cubic B-spline basis functions, with a simple roughness based penalty on its second derivative. Afterwards, using the estimated curve $\hat{\lambda}'(t)$ and an input function $u(t)$ (a step function at the current change-point candidate), we estimate $\hat{\gamma}^{(0)}$ with a simple linear regression procedure. These estimates are used to update the $\hat{\beta}$ estimate using $\mathcal{L}_{\hat{\gamma}^{(0)}}$ as a regularization penalty. The process of estimating $\hat{\gamma}$ is alternatively repeated with the procedure of estimating $\hat{\beta}$, until convergence is found.

Figure 7.2 shows the rate function's initial first derivative estimate, taken from the smoothed estimate $\hat{\lambda}$ shown in Figure 7.1, and the true first derivative (the dashed line).

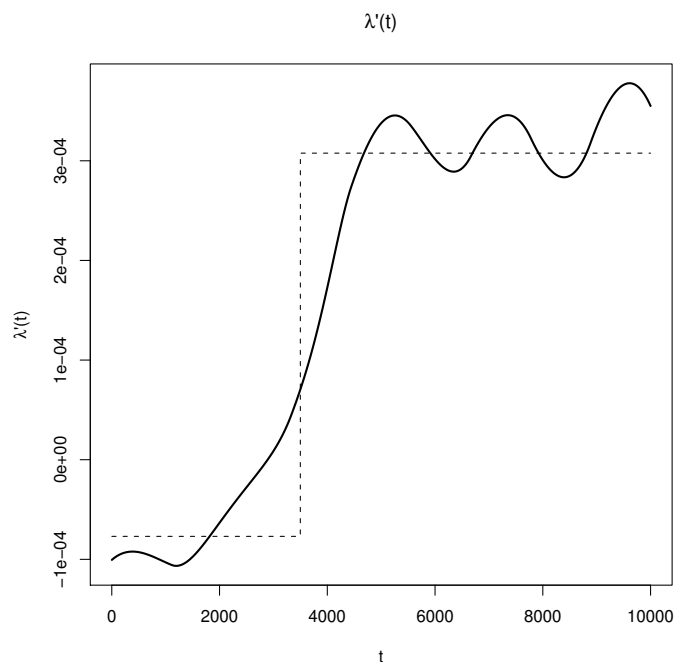


Figura 7.2: The function $\hat{\lambda}'(t)$, estimated with a simple roughness smoothing (solid line) and the true first derivative (dashed line)

In Figure 7.3 we have the true rate function $\lambda(t)$. The colored curves represent estimates penalized by the L operator, each one at a given change point candidate, as specified by a color scale rule. The candidates tested in this case are 23 points $c_j = \{416, 833, 1250, \dots, 9583\}$, equally spaced in the interval $[0, 10000]$, and since the true change point is $c = 3500$, the nearest candidate is given by $c_8 = 3333$ ($c_9 = 3750$ being also a likely choice). The penalization constant here is $\xi = 10^6$, and the red estimate corresponds to the one penalized by the correct change point candidate. The color rule goes from red (the true one) to blue (corresponding to the farthest change point candidates, at each direction). The true rate function is plotted in black.

Figure 7.4 shows the AIC-like criterion (actually simply the negative log likelihood, since the number of parameters remains the same) for each of the estimatives, as a function of the change point candidate, with $\xi = 10^6$. We used this AIC-like criterion based on the procedure that was proposed by Dias and Garcia (2007), even though formally our context is different. One can see in Figure 7.4 that the true candidate

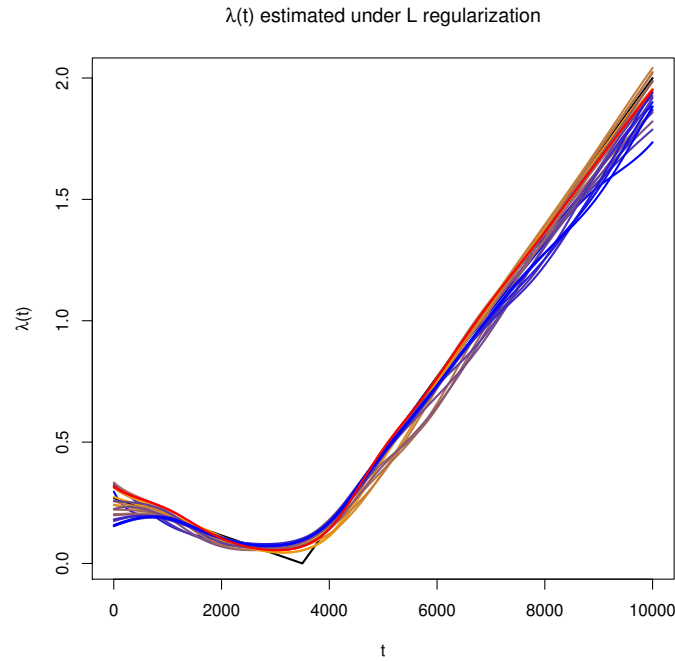


Figure 7.3: True function $\lambda(t)$ and its estimated obtained with the L operator, $\xi = 10^6$

almost minimizes the negative log likelihood – though a neighbour is the actual observed minimum – and the underlying distribution looks convex despite noise.

Increasing the tuning parameter to $\xi = 10^7$ results in a sharpening of the differences among the curves. The one based on the correct change point (still red) remains a good estimator of the rate function, but those based on the farthest change points have deteriorated in quality already. We now propose that it is possible to find the best estimate based on an AIC-like criterion.

Figure 7.3 clearly shows the estimates – because there is not much differences between them – while Figure 7.5 is much harder to analyze. So we plotted in Figure 7.6 each of the estimates, grouped in clusters of three, with each change point candidate marked by an arrow of corresponding color. The estimate corresponding to the change point c_8 , which is the nearest one to the true change point $c = 3500$, is found at subplot (c), colored red.

Figure 7.7 shows the same results as Figure 7.4, except that now the tuning

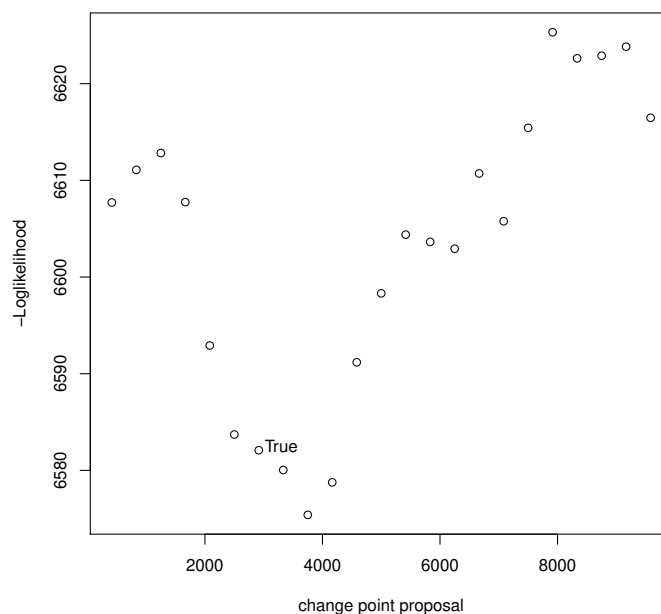


Figura 7.4: Negative log likelihood for the L penalized estimates, as a function of c , $\xi = 10^6$

parameter is $\xi = 10^{10}$. Again the minimum is not reached by the true candidate, though the difference between the actual minimum and the observed minimum seems not to be statistically significant. Referring back to Figure 7.6, we can see that the change point candidate that minimizes the negative log likelihood, given by $c_{10} = 4166$, is indeed a very good fit to the original rate function ((d), the curve colored in blue) despite its change point candidate being a bit too far from the true value.

7.4 Mexico City Ozone Pollution Data

Air pollution is a well-known health hazard, frequently associated with respiratory and cardiovascular diseases, particularly in developing countries where environmental policies are not widespread (see, for example, Achcar et al., 2008; Achcar et al., 2008; Gouveia et al., 2006; Braga et al., 2002). Among the pollutants that affect large cities is ozone gas, when its concentration remains above a certain threshold (as detailed in

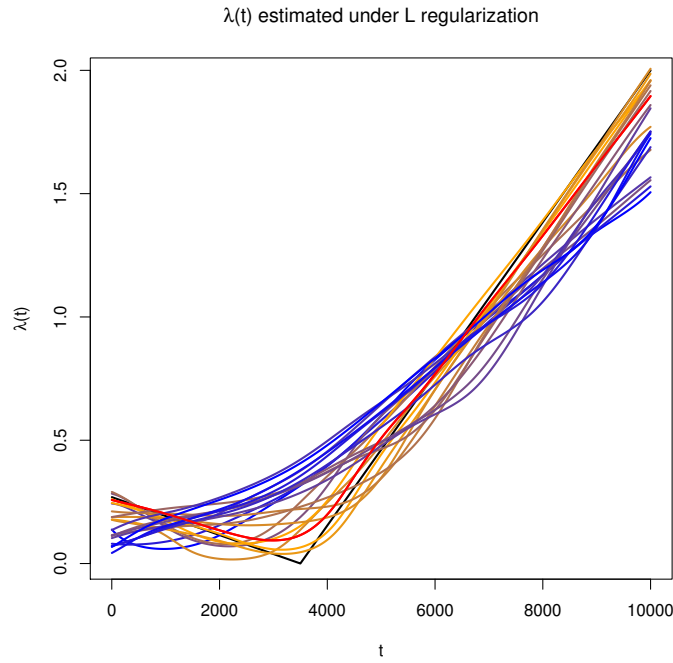


Figure 7.5: True function $\lambda(t)$ and its estimated obtained with the L operator, $\xi = 10^7$

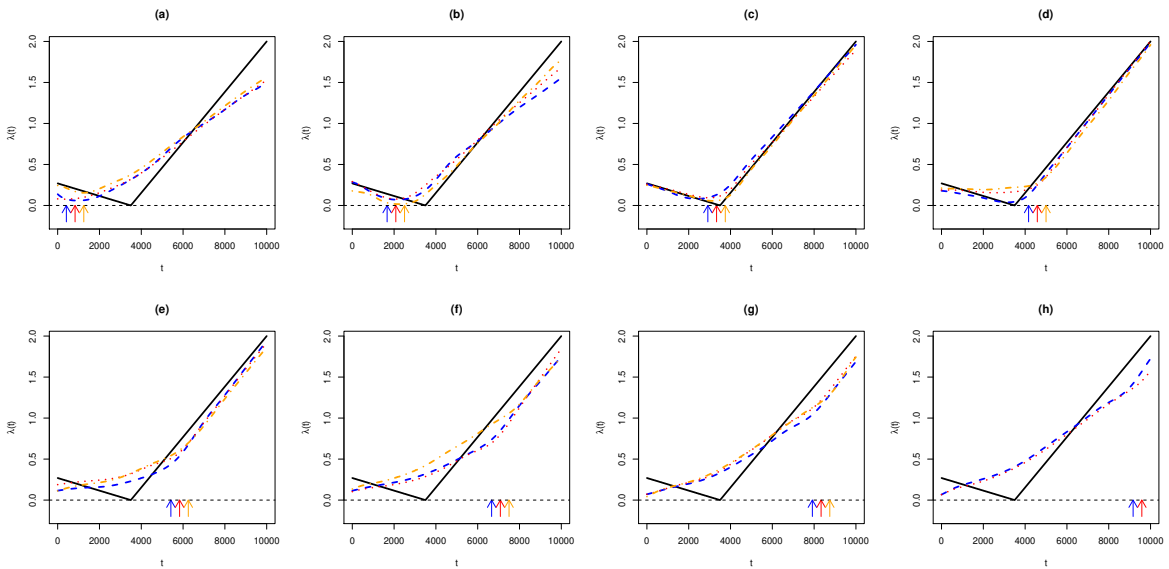


Figure 7.6: True function $\lambda(t)$ and its estimated obtained with the L operator, $\xi = 10^7$ – (a) shows candidates c_1 , c_2 and c_3 , with their corresponding positions marked by arrows, and so on for each subplot

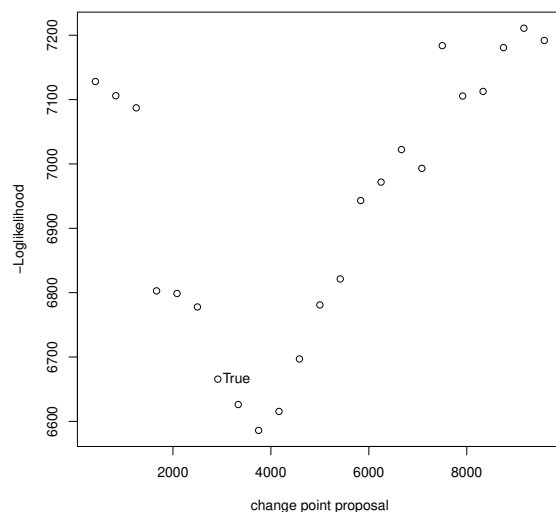


Figura 7.7: Negative log likelihood for the L penalized estimates, as a function of c , $\xi = 10^8$

research by Bell et al., 2004).

Counting every time the ozone concentration goes above a threshold can be regarded as a Poisson process, with a certain rate function $\lambda(t)$. Such model was proposed in Achcar, Fernández-Bremauntz and Rodrigues (2008) for the ozone concentration data in the troposphere of Mexico City. The rate function $\lambda(t)$ is assumed to be nonhomogeneous, since it varies as population and industrialization increases – in a sometimes spurious positive correlation with time. Furthermore, assessing the impact of public health intervention measures, such as new laws regarding the emission of pollutants, is also a research interest.

Ozone (O_3) is formed by chemical reactions between NO_x and volatile organic compounds (VOCs) in the presence of sunlight. It can be considered “good” or “bad”, depending on its location in the atmosphere. The “good” ozone occurs naturally in the stratosphere, about 10 to 30 miles above the Earth’s surface and forms a layer that protects life on Earth from harmful sun rays. Tropospheric ozone is considered “bad”, and it can be formed by the reaction of gases emitted from vehicle exhausts, industrial emissions, gasoline vapors, solvents and chemicals and natural sources. The chemical

reactions involved in the formation of tropospheric ozone are a complex series of cycles in which carbon monoxide and volatile organic compounds are oxidized to water vapor and carbon dioxide through chemical and photochemical reactions. Tropospheric ozone can cause various health problems in humans and animals such as chest pain, cough, and throat irritation, and is also harmful to plantlife.

The O_3 pollutant data from Mexico City was collected by the Monitoring Network of Mexico City (www.sma.df.gob.mx/simat/) and represents eighteen years of measurements (from January 1, 1990 to December 31, 2008) of the overall daily maximum measurements for the city. The measurements are obtained minute by minute and the average hourly result is reported at each station. The NE data is obtained by taking the daily maximum among the average values reported by all monitoring stations in the city (Achcar et al. 2009). Even though the Mexican standard for ozone is 0.11 parts per million (0.11 ppm) and 0.22 ppm is the threshold used in Mexico City for declaring an emergency situation, we are going to consider the value 0.17 ppm as the threshold to indicate that an ozone peak has occurred. The main reason for this is that the threshold 0.11ppm is surpassed often and the threshold 0.22 ppm is rarely surpassed – hence the choice of an intermediate value. Another reason is that the environmental authorities in Mexico City have been thinking of reducing the value of the threshold used to declare emergency situations (Achcar et al., 2011). Any other value could be used, specially if the interest is either in obtaining information about the behaviour of the pollutant when a certain threshold is considered or in seeing how this behaviour could be affected by government decisions(Achcar et al., 2010).

In the analysis of air pollution data, various models and methodologies have been presented in the literature: the use of stochastic volatility models, proposed by Achcar, Rodrigues, Saldiva and Zozolotto (2008); homogeneous Poisson processes, by Raftery (1989); nonhomogeneous Poisson Process, by Achcar et al. (2008) and Achcar et al. (2009); and finally, the use of homogeneous Poisson processes in the presence of change points, by Achcar et al. (2009). Such methodologies were either based on Bayesian methods or nonparametric empirical distribution functions.

Some authors have already addressed the problem of the estimation of change points by functional data analysis. Grégoir and Zouhir (2002) estimated change points by local linear smoothing regression. Prieur (2007) estimated just a single change point with local linear smoothing under a condition of linear dependence to another covariable.

Other author as Wang (1999) estimate change-points by the wavelet transformation.

In this section we apply our proposed models to the pollution data from Mexico City. As an example of the type of real world results that one may obtain from the analysis described in the present work, we consider the data from the NE region. This region is important because the wind blows to the SW into the rest of Mexico City, and also because some of the ozone precursors are in fact produced in this region (Achcar et al., 2008). As explained before, these data correspond to the daily average measurements during the period from January 1st, 1990, to December 31st, 2008, and we considered only the days on which the daily average exceeded the 0.17 ppm threshold, with a total of 1541 observations.

Figure 7.8 shows the estimated rate $\lambda(t)$ for the Mexico City pollution data, using different smoothing parameters. The black line represents the maximum likelihood estimate of ordinary exponentiated splines. The smoothed splines estimate with $\xi = 10^5$ is almost the same as the maximum likelihood and is shown (if visible at all) in blue. The orange line shows a smoothed estimate with $\xi = 10^{10}$, and the red line shows the smoothed estimate with $\xi = 10^{15}$. In all of the scenarios, we used twelve basis functions with equispaced knots on the interval $[0, 6935]$. The mean value function $m(t)$ estimates are obtained by integrating the estimates $\lambda(s)$ over the interval $[0, t]$, that is,

$$\hat{m}(t) := \int_0^t \hat{\lambda}(s) ds,$$

and they are shown in Figure 7.9, along with the empirical $\hat{m}(t)$. We must observe that, while the fit for the $m(t)$ curves seems very good while the smoothing parameter is low; in our case, 10^{15} is clearly too much smoothing – but examining differences in $\lambda(t)$ estimation should be our focus, since it represents a relative difference.

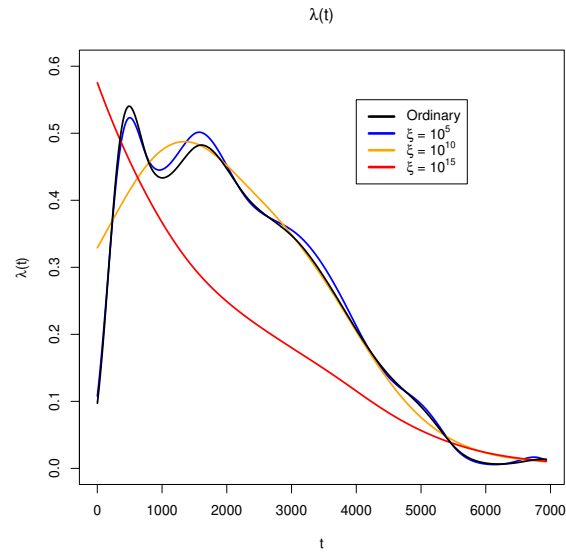


Figure 7.8: Estimated λ for the Mexico data. The black line is the maximum likelihood estimate; the blue one is the smoothed splines estimate with $\xi = 10^5$; the orange line shows a smoothed estimate with $\xi = 10^{10}$; the red line shows the smoothed estimate with $\xi = 10^{15}$.

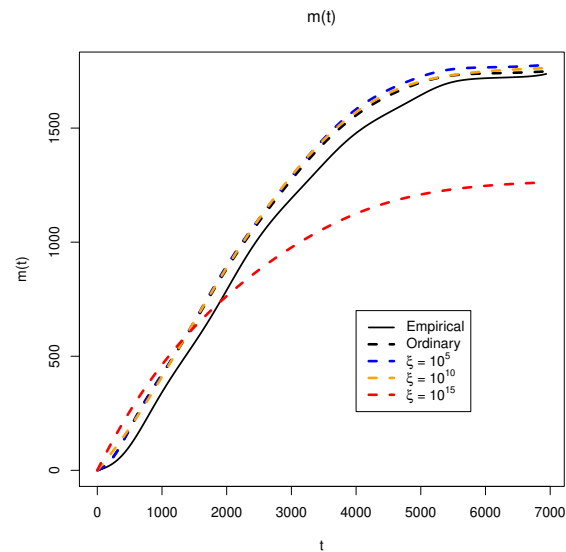


Figure 7.9: Empirical and estimated mean value function $m(t)$ for the Mexico pollution data. Black line is the maximum likelihood estimate; the orange line shows a smoothed estimate with $\xi = 10^{10}$; the red line shows the smoothed estimate with $\xi = 10^{15}$.

Figure 7.10 shows the estimates when the penalization based on the discontinuous first derivative L_γ is used. We used an equispaced mesh of change point candidates composed of 40 points, and the same tuning parameter of 10^5 . Inspecting the final estimates $\hat{\beta}$, we notice that this choice of tuning parameter gives a penalization ranging from 4.3 to 19, whereas the likelihood is about 3600, clearly dominating the fit criterion. Consequently, the estimates are not very different, as observed. Figure 7.11 shows that the minimum negative log likelihood is achieved at day (520), out of a period of 6935 days, suggesting an early change point in the second year of measurements.

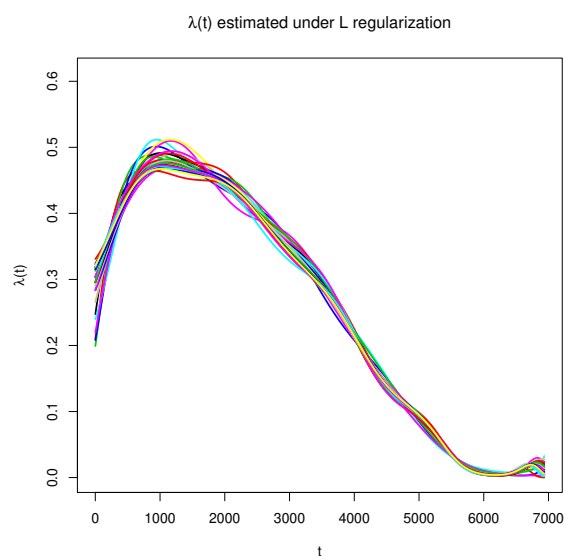


Figure 7.10: The estimated rate functions $\lambda(t)$ when a penalty based on a change point candidate is applied. The tuning parameter ξ is equal to 10^5 .

The procedure was repeated, now using a tuning parameter $\xi = 10^6$. The results are shown, overplotted, at Figure 7.12. Inspecting the penalization criterion for the final estimates $\hat{\beta}$, we notice the penalization criterion now ranges from 13.3 to 127, while the likelihood still has values around 3600. Figure 7.13 now has a minimum on day 866, with day 693 being a close candidate. This does not contradict the result from Figure 7.11, that the most significant change point happened during the second year. Since the larger tuning parameter did sharpen the differences of the final estimates, we provide a better plot in Figure 7.14, where pairs of curves are plotted (though only a

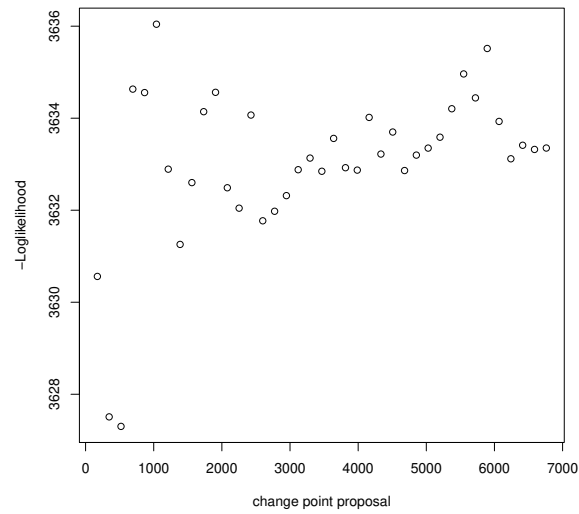


Figura 7.11: Negative log likelihood for each change point candidate, when $\xi = 10^5$.

subset of the 40) with the corresponding change point candidate marked by an arrow.

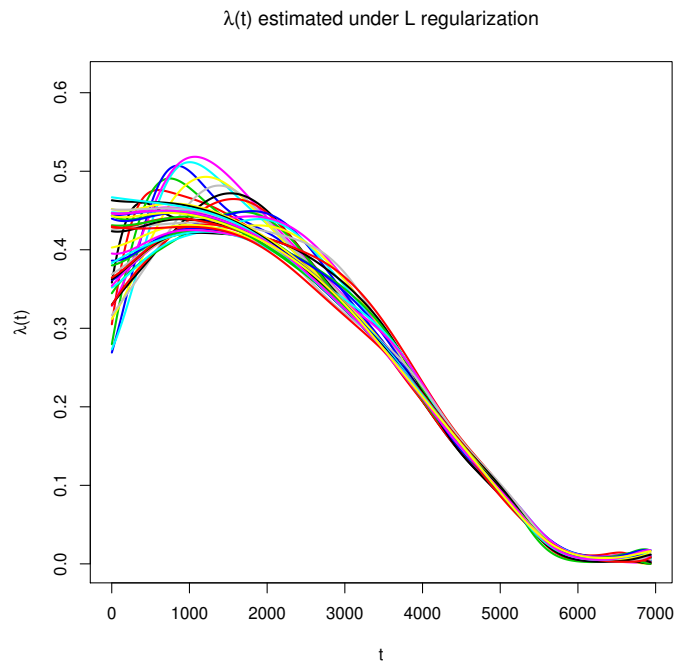


Figura 7.12: The estimated rate functions $\lambda(t)$ when a penalty based on a change point candidate is applied. The tuning parameter ξ is equal to 10^6 .

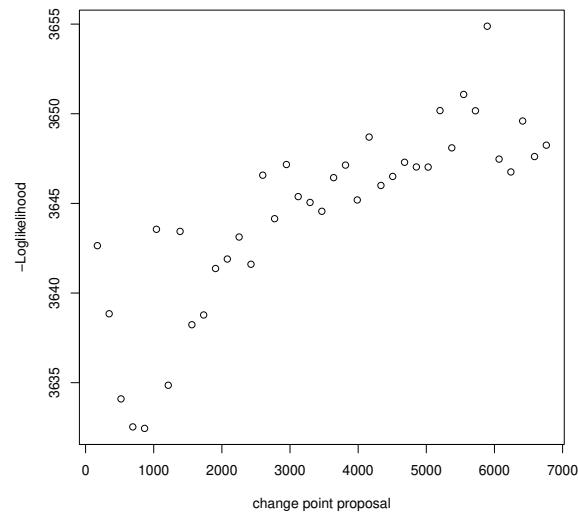


Figure 7.13: Negative log likelihood for each change point candidate, when $\xi = 10^6$.

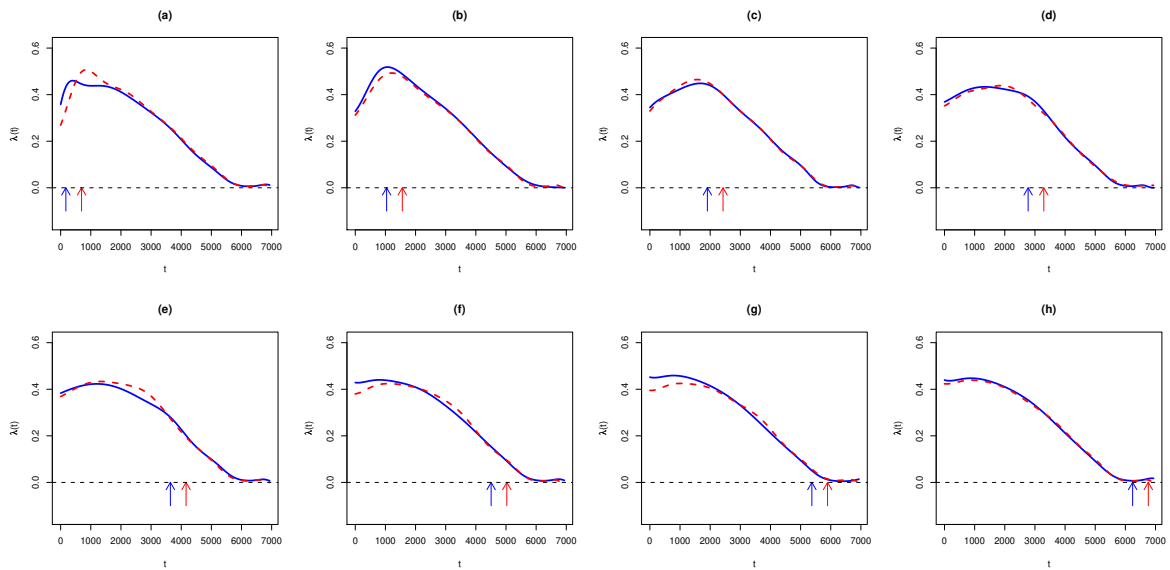


Figure 7.14: Different estimates with penalizations based on a change point candidate – marked by an arrow – with a penalization parameter of 10^6 .

We can use two change point candidates with a small modification of the algorithm. We will now search in a lattice of candidates (c_i, c_j) , $c_j > c_i$. Instead of looking at the 39 equally spaced candidates, we choose to test only 19 candidates, yielding a lattice

of 171 pairs. Figure 7.15 shows the negative log likelihood for each pair of change points candidates. The first 18 pairs, colored black, corresponds to the pairs $(1, 2), (1, 3), \dots, (1, 19)$. The next 17, colored red, corresponds to $(2, 3), (2, 4), \dots, (2, 19)$, and so on, up to pair $(18, 19)$. The arrows in Figure 7.15 indicate for which pairs the estimated rate function λ were plotted in Figure 7.16. Again the change points near the start of the measurement period are the most likely candidates, with the pair $(173, 346)$ yielding the lowest negative log likelihood, and $(173, 520)$ also being a good choice.

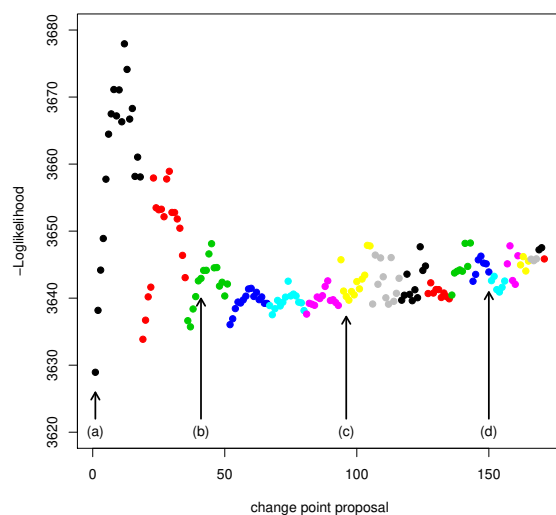


Figure 7.15: AIC-like criterion for each pair of change points candidates.

For illustration purposes, Figure 7.16 shows 4 of 171 rate function estimates with their corresponding change point candidates marked by the arrows on the t -axis.

We did not formally compare the model with a single change point against the model with two change points, nor the models with change points against the smoothed spline model. This could be done by considering the proper number of parameters as in the canonical AIC formulation. However, we are not sure whether ξ , given our exploratory way of finding it, counts as an estimated parameter or not. This does not make a difference when the number of change point candidates is the same, because we can simply compare their likelihood, but finding the correct constant and adequate true

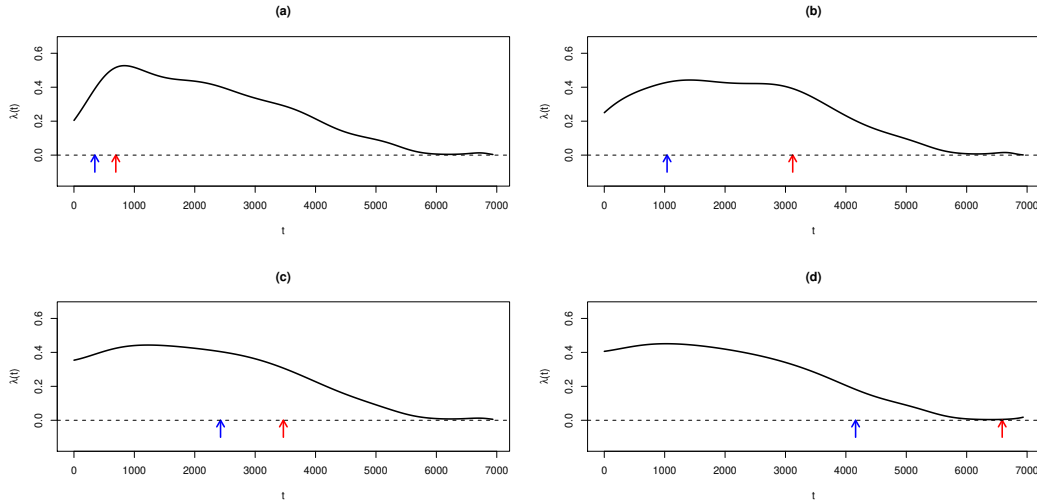


Figura 7.16: Rate function estimates with change point candidates corresponding to the pairs marked by arrows at Figure 7.15

number of parameters when the models differ is not straightforward. Also, one must keep in mind that the penalization norms, related to either the roughness or the L_γ operator, are not included in the AIC formulation, and thus we would expect larger AIC values for them. This hampers the functionality of the AIC as a decision rule. Given this word of caution, we computed the AIC as

$$AIC_{\mathcal{L}} = 2(K + \#\|\gamma\| + 1) - 2 \log \mathcal{L}(\hat{\beta}, L_{\hat{\gamma}} \mid t_1, \dots, t_n, T),$$

where $\#\|\gamma\|$ represents the number of change points (0 for the roughness smoothing procedure) and +1 is due the tuning parameter ξ . The resulting values are shown in Table 7.1.

Tabela 7.1: AIC for the different models.

Model	AIC
Ordinary Spline	7272.716
Smoothing Spline, $\xi = 10^{10}$	7272.503
L_γ penalized Spline, best fit, 1 change point, $\xi = 10^5$	7282.603
L_γ penalized Spline, worst fit, 1 change point, $\xi = 10^5$	7300.086
L_γ penalized Spline, best fit, 1 change point, $\xi = 10^6$	7292.902
L_γ penalized Spline, worst fit, 1 change point, $\xi = 10^6$	7337.765
L_γ penalized Spline, best fit, 2 change points	7287.885
L_γ penalized Spline, worst fit, 2 change points	7385.886

7.5 Conclusions

Our methods try to expand the spline-based theory for Poisson processes, relaxing the smoothness hypothesis regarding the true function $\lambda(t)$ by including a penalization that aims to capture a derivative discontinuity, while still smoothing the function elsewhere. The final estimate still is a cubic spline function. Although it remains an open problem to find an algorithm for optimal smoothness parameter, we have shown that these methodology can detect the true change point regardless of the tuning parameter choice.

In the application to the real dataset we found that, while the Mexico City pollution data is definitely from a nonhomogeneous process, the sharpest changes occur only in the first or second year and every subsequent change was smooth enough to be captured by the spline structure alone.

Acknowledgements

We would like to thank Professor Ronaldo Dias from the University of Campinas for his insightful comments on the smoothing spline problems, and John Davis from the University of Wisconsin-Madison for proofreading this paper and giving additional suggestions. We also would like to thank the support from CAPES and CNPq agencies, and the EPIFISMA laboratory.

Referências Bibliográficas

- Achcar, J. A.; Rodrigues, E.R.; Saldiva, P.H.N. and Zozolotto, H.C. – Bivariate stochastic volatility models applied to Mexico City ozone pollution data, In Gaia C. Romano and Alice G. Conti. (Org.). *Air Quality in the 21st Century*, 1st ed., New York: Nova Publishers, v. 01, p. 0001–0030, 2008.
- Achcar, J. A.; Fernández-Bremauntz, A. A. and Rodrigues, E.R. – Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model, *Environmetrics*, v. 19, Issue 5, p. 469–485, 2008.
- Achcar, J. A.; Rodrigues, E.R.; and Tzintzun, G. – Using stochastic volatility models to analyse weekly ozone averages in Mexico City, *Environmental and Ecological Statistics*, 2010.
- Achcar, J. A.; Rodrigues, E.R.; and Tzintzun, G. – Modelling inter-occurrence times between ozone peaks in Mexico city in the presence of multiple change-points, *Brazilian Journal of Probability and Statistics*, v. 25, No. 2, p. 183-204, 2011.
- Achcar, J. A.; Rodrigues, E.R.; Paulino, C. D. and Soares, P. – *Non-homogeneous Poisson models with a change-point: an application to ozone peaks in Mexico City*, p. 001–029, 2009.
- Bell, M. L.; McDermontt, A.; Zeger, S. L.; Samet, J. M. and Dominici, F. – Ozone and short-term mortality in 95 US urban communities 1987 – 2000, *Journal of the American Medical Society*, v. 292, p. 2372–2378, 2004.

- de Boor, C. – *A Practical Guide to Splines, Revised Edition*, New York: Springer, 2001.
- Braga, A.L.F.; Zanobetti, A. and Schwartz, J. – The effect of weather on respiratory and cardiovascular deaths in 12 U.S. Cities, *Environmental Health Perspectives*, v. 110, p. 859–863, 2002.
- Dias, R. and Garcia, N. – Consistent estimator for basis selection based on a proxy of the Kullback-Leibler distance, *Journal of Econometrics*, v. 141, No. 1, p. 167–178, 2007.
- Ferraty, F. and Vieu, P. – *Nonparametric Functional Data Analysis*, New York: Springer, 2006.
- Gouveia, N.; Freitas, C.U.; Martins, L.C.; Marcilio, I.O. – Hospitalizações por causas respiratórias associadas à contaminação atmosférica no município de São Paulo, Brasil, *Cadernos de Saúde Pública*, Rio de Janeiro, v. 22, p. 2669–2677, 2006.
- Grégoire, G. and Zouhir, H. – Change point estimation by local linear smoothing, *Journal of Multivariate Analysis*, v. 83, No. 1, p. 56–83, 2002.
- Leemis, L. M. – Nonparametric Estimation of the cumulative intensity function for a nonhomogeneous Poisson process, *Management Science*, v. 37, No. 7, p. 886–900, 1991.
- Lyche, T. and Schumaker, L. L. – *L-Spline Wavelets, Wavelets: Theory, Algorithms, and Applications*, p.197–212, 1994.
- Prieur, C. – Change point estimation by local linear smoothing under a weak dependence condition, *Mathematical Methods of Statistics*, v. 16, p. 25–41, 2007.
- Raftery, A. E. – Are the ozone exceedance rate decreasing? Comment on the paper “Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone” by R.L.Smith, *Statistical Science*, Hayward, v.4, p.378-381, 1989.

Ramsay, J. O. –Differential equation models for statistical functions, *The Canadian Journal of Statistics*, v. 28, No. 2, p. 225–240, 2000.

Ramsay, J. O.; Hooker, G.; Campbell, D. and Cao, J. – Parameter estimation for differential equations: A generalized smoothing approach, *Journal of the Royal Statistical Society*, Part 5, p. 741-796, 2007.

Ramsay, J. O. and Silverman, B. W. – *Functional data analysis, Second Edition*, New York: Springer-Verlag, 2005.

Wahba, G. – *Spline models for observational data*, Philadelphia: Society for Industrial and Applied Mathematics, 1990.

Wang, Y. – Change-point analysis via wavelets for indirect data, *Statistica Sinica*, p. 103-117, 1999.

Yaglom, A. M. – *Correlation theory of stationary and related random functions - volume I*, New York: Springer-Verlag, 1987.

.1 Estimating Differential Equation Parameters from Data

The paper by Ramsay et al. (2007) proposes the modelling of differential equations' parameters of $Y(t)$ from observed data. In order to do this, they considered a linear differential equation of the form

$$y^{(p)}(t) = -\gamma_0 y(t) - \gamma_1 y'(t) - \dots - \gamma_{(p-1)} y^{(p-1)}(t) + u(t), \quad (8)$$

which can be written as

$$L_\gamma y(t) = \gamma_0 y(t) + \gamma_1 y'(t) + \dots + \gamma_{(p-1)} y^{(p-1)}(t) + y^{(p)}(t) - u(t) = 0 \quad (9)$$

where $\gamma_0, \gamma_1, \dots, \gamma_{p-1}$ are the parameters of the differential equation to be estimated, and L_γ is an operator that annihilates the differential equation structure (in the sense that $L_\gamma \hat{y}(t) = 0$ should $\hat{y}(t)$ be a solution of (8)). Suppose that the function $y(t)$ is observed at the moments $t = t_1, t_2, \dots, t_n$, giving us the observations y_1, y_2, \dots, y_n . A data fitting criterion, supposing that the observation errors have distribution g , is the negative log likelihood, given by the equation

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\gamma}) = -\ln g(\mathbf{e} \mid \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\gamma}),$$

where

$$e_i = y_i - \boldsymbol{\beta}(\boldsymbol{\sigma}, \boldsymbol{\gamma}; \boldsymbol{\xi})' B(t_i),$$

and g is the distribution of errors and $\boldsymbol{\sigma}$ their covariance. The closer a function y is from a solution of the differential equation (8), the lower the penalty

$$\text{PEN}_L(y) = \int [L_\gamma(y(t))]^2 dt \quad (10)$$

is, being zero when the function y is in fact a solution of such differential equation. The integration is over the interval in which we observe the process.

To find the estimate $\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}; \boldsymbol{\xi})$, one combines the data fitting criterion (8) with the differential equation fidelity criterion (10) on a single criterion: the penalized log likelihood, given by,

$$\mathcal{L}_\xi(\boldsymbol{\beta} \mid \boldsymbol{\sigma}, \boldsymbol{\gamma}) = -\ln g(\mathbf{e} \mid \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\xi}) + \text{PEN}_L(y \mid \boldsymbol{\xi}).$$

The estimation of the coefficients $\boldsymbol{\gamma}$ begins with a rough estimate $\hat{y}^{(0)}(t)$, smoothed up to its $p + 2$ derivate (so that the p -th derivative is smooth which requires order $p + 4$ B-splines basis functions). The concurrent linear model procedure consists into using an operator L to define a new roughness penalty based on $\boldsymbol{\gamma}^{(0)}$ estimated from this initial fit $\hat{y}^{(0)}(t)$ and its derivatives, by simple linear regression. These steps will be

alternated until we have convergence of both the parameters γ and the estimate $\hat{y}(t)$.

Apêndice A

Análises adicionais dos capítulos: 5 e 6

A.1 Análise de sensibilidade de modelos com pontos de mudança

Dentre as distribuições *a priori* testadas para o parâmetro k_i estão: Uniforme(0,T), Uniforme(0,1), Gama(0.001,0.001), Gama(100,100), além de outros valores para os hiperparâmetros destas distribuições. Foi também testado a transformação $\log(k)$, na qual $\log(K) \sim N(a_1, b_1)$, com vários valores dos hiperparâmetros a_1, b_1 , entre eles, N(0;0.01), N(10;10), N(0;1). Na maioria dos testes a convergência das simulações da cadeia de Markov não foi satisfatória. O melhor resultado para este modelo foi obtido utilizando uma distribuição exponencial truncada (a_n, g) , onde foram testadas valores de g de 3 a 20. O que apresentou melhor resultado foi a distribuição exponencial truncada com parâmetro da média igual a 0.95 e truncada em 3 em alguns modelos e em outros com parâmetro da média 0.99 e truncada em 6. Nos parâmetros $\beta_1, \beta_2, \alpha_1$ e α_2 , também foi realizado uma análise de sensibilidade, na qual alguns dos valores propostos dos hiperparâmetros para β_1 foram: $\Gamma(0.001, 0.001)$, $\Gamma(0.01 * 100, 100)$, $\Gamma(0.01 * 10, 10)$, $\Gamma(0.001 * 1000, 1000)$ entre outros. Para o parâmetro β_2 foram os seguintes valores dos

hiperparâmetros: $\Gamma(0.001, 0.001)$, $\Gamma(0.03*100, 100)$, $\Gamma(0.003*100, 100)$, $\Gamma(0.01*100, 100)$ entre outros. Para os parâmetros α_1 e α_2 , testamos o primeiro conjunto de prioris não informativas, propostas por (Achcar, et al., 1998), num segundo momento foram testadas distribuições $U(0,T)$, $U(0,1)$ entre outros valores dos hiperparâmetros. Num terceiro momento testamos distribuições *a priori* $\Gamma(0.001, 0.001)$, $\Gamma(2*1, 1)$, $\Gamma(1, 1)$ entre outros valores. A escolha das distribuições *a priori* dos parâmetros foi feita de tal forma que algumas distribuições *a priori* pouco informativas e outras não informativas.

A.2 Figuras dados simulados, modelos sem pontos de mudança de mudança

Na Figura (A.1), mostramos a convergência das cadeias para os parâmetros do Modelo I sem pontos de mudança, para os dados simulados.

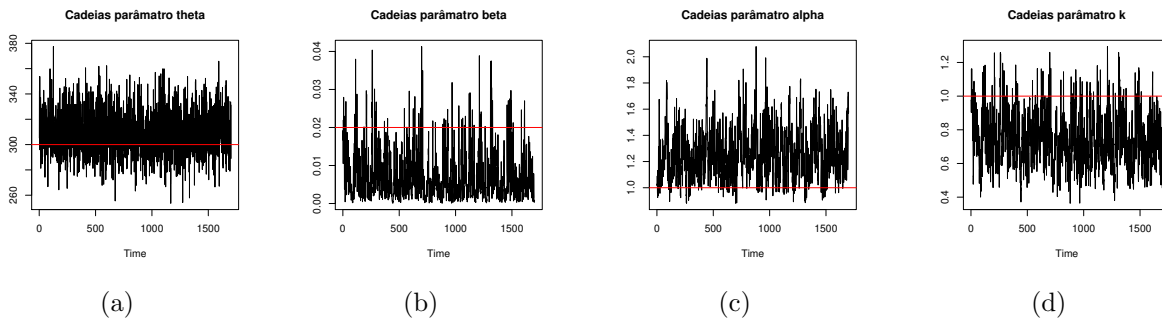


Figura A.1: Convergência das cadeias para os parâmetros do Modelo I sem pontos de mudança, sendo que o traço na horizontal é o valor verdadeiro de cada parâmetro.

Para verificar convergência, as distribuições *a posteriori* dos parâmetros foram estimadas partindo de dois pontos iniciais diferentes. De uma forma geral, todas as cadeias convergiram ao seu valor verdadeiro. Aqui apresentamos apenas um dos resultados.

Na Figura (A.2), mostramos as distribuições *a posteriori*, *a priori* e o valor verdadeiro de cada parâmetro, para o Modelo I sem pontos de mudança, para os dados simulados.

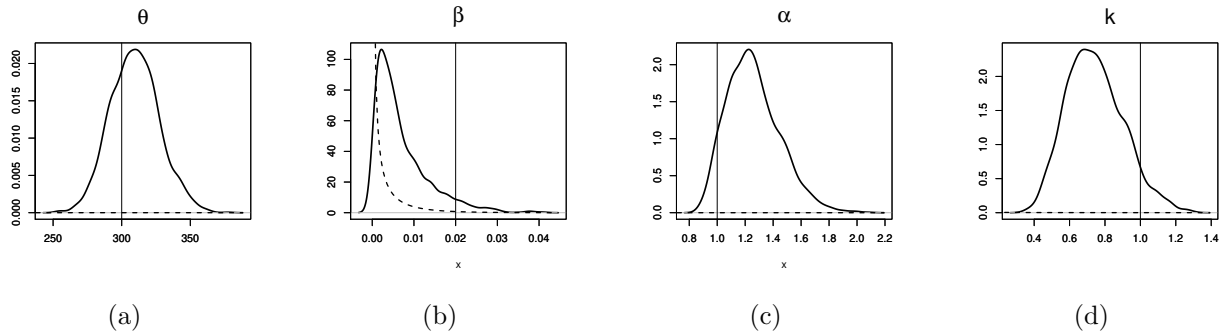


Figura A.2: Distribuição *a posteriori* dos parâmetros referente ao Modelo I, sem pontos de mudança. A distribuição *a posteriori* corresponde a linha cheia e a distribuição *a priori* a linha tracejada. O verdadeiro valor de cada parâmetro é representado pela linha vertical.

Na Figura (A.3), mostramos a convergência das cadeias para os parâmetros do Modelo II, sem pontos de mudança, para os dados simulados.

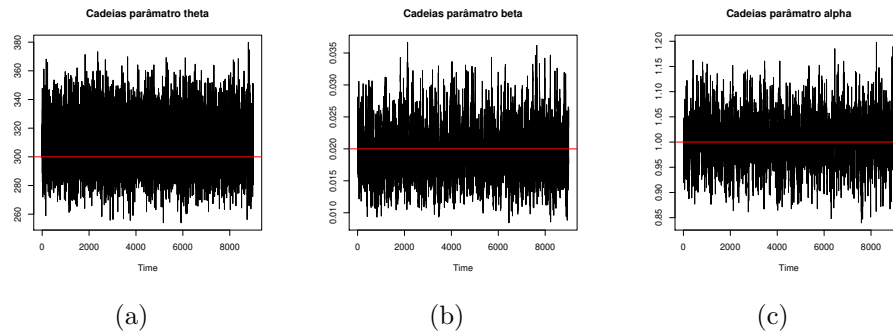


Figura A.3: Convergência das cadeias para os parâmetros do Modelo II sem pontos de mudança, sendo que o traço na horizontal é o valor verdadeiro de cada parâmetro.

Na Figura (A.4), mostramos as distribuições *a posteriori*, *a priori* e o valor verdadeiro de cada parâmetro, para o Modelo II, sem pontos de mudança, para os dados simulados.

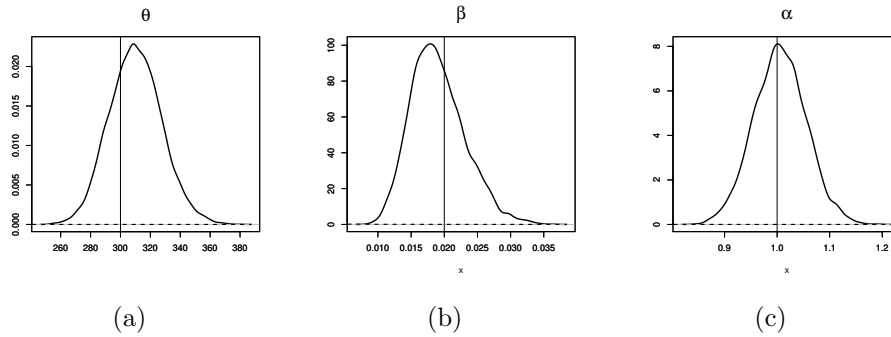


Figura A.4: Distribuição *a posteriori* dos parâmetros referente ao Modelo II, sem pontos de mudança. A distribuição *a posteriori* corresponde a linha cheia e a distribuição *a priori* a linha tracejada. O verdadeiro valor de cada parâmetro é representado pela linha vertical.

Na Figura (A.5), mostramos a convergência das cadeias para os parâmetros do Modelo III, sem pontos de mudança, para os dados simulados.

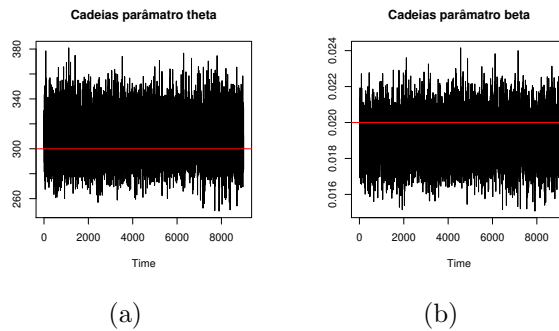


Figura A.5: Convergência das cadeias para os parâmetros do Modelo III, sem pontos de mudança, sendo que o traço na horizontal é o valor verdadeiro de cada parâmetro.

Na Figura (A.6), mostramos as distribuições *a posteriori*, *a priori* e o valor verdadeiro de cada parâmetro, para o Modelo III, sem pontos de mudança, para os dados simulados.

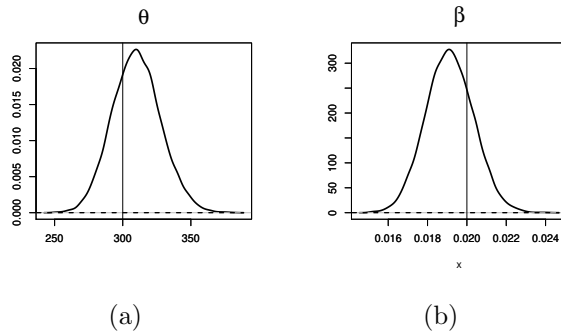


Figura A.6: Distribuição *a posteriori* dos parâmetros referente ao Modelo III, sem pontos de mudança. A distribuição *a posteriori* corresponde a linha cheia e a distribuição *a priori* a linha tracejada. O verdadeiro valor de cada parâmetro é representado pela linha vertical.

A.3 Figuras dados simulados, modelos com um ponto de mudança de mudança

Na Figura (A.7), mostramos a convergência das cadeias para os parâmetros do modelo I, com um ponto de mudança, para os dados simulados.

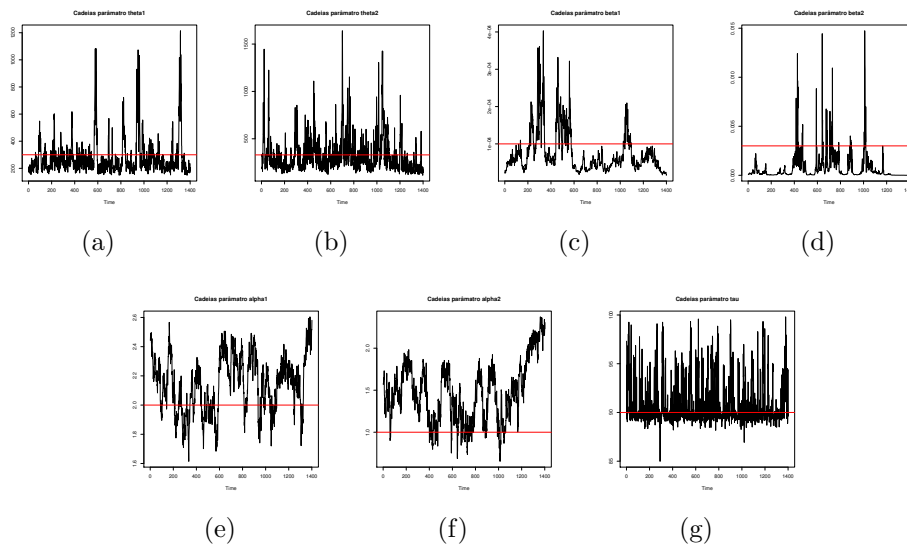


Figura A.7: Convergência das cadeias para os parâmetros do modelo II, com um ponto de mudança, sendo que o traço na horizontal é o valor verdadeiro de cada parâmetro.

Na Figura (A.8), mostramos as distribuições *a posteriori*, *a priori* e o valor verdadeiro de cada parâmetro, para o modelo II, com um ponto de mudança, para os dados simulados.

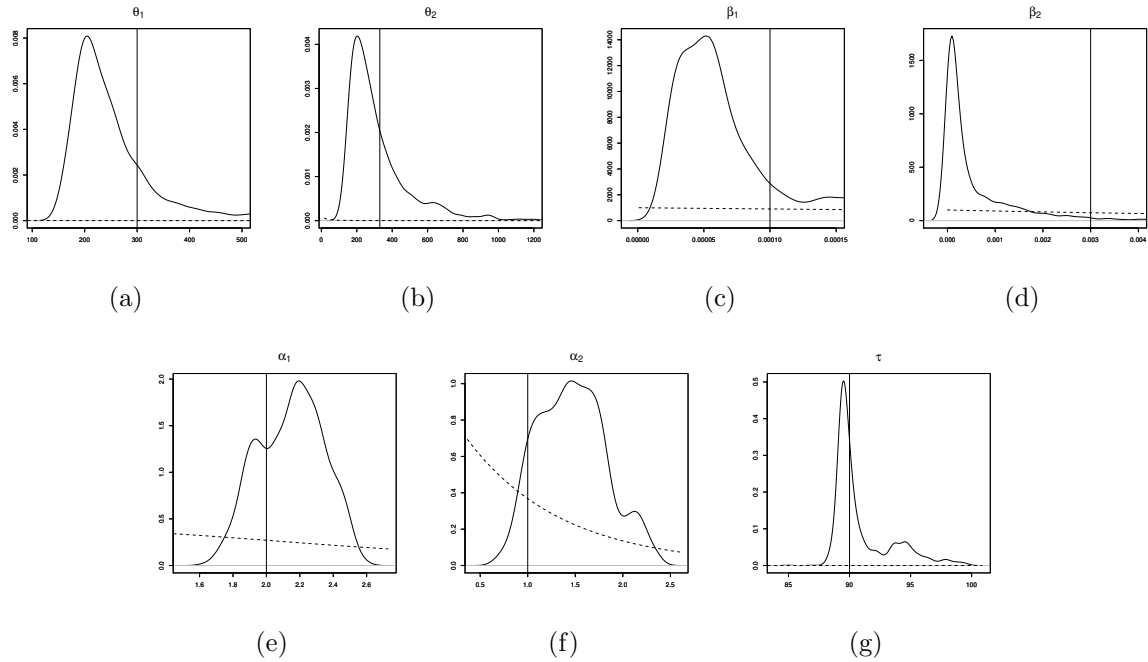


Figura A.8: Distribuição *a posteriori* dos parâmetros referente ao modelo II, com um ponto de mudança. A distribuição *a posteriori* corresponde a linha cheia e a distribuição *a priori* a linha tracejada. O verdadeiro valor de cada parâmetro é representado pela linha vertical.

Na Figura (A.9), mostramos a convergência das cadeias para os parâmetros do modelo III, com um ponto de mudança, para os dados simulados.

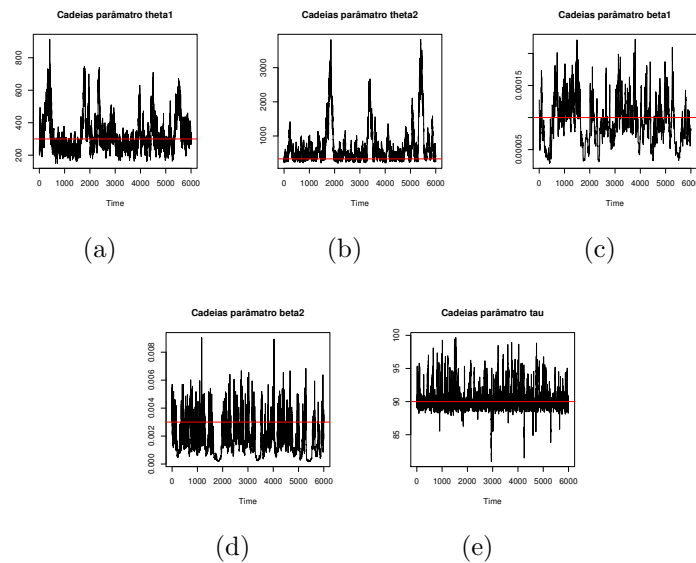


Figura A.9: Convergência das cadeias para os parâmetros do modelo III, com um ponto de mudança, sendo que o traço na horizontal é o valor verdadeiro de cada parâmetro.

Na Figura (A.10), mostramos as distribuições *a posteriori*, *a priori* e o valor verdadeiro de cada parâmetro, para o modelo III, com um ponto de mudança, para os dados simulados.

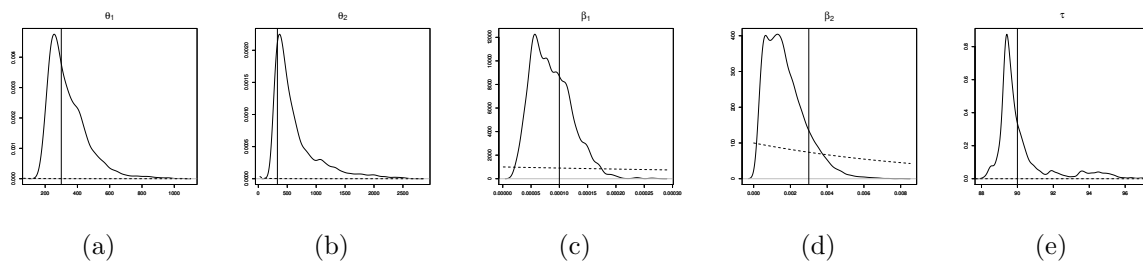


Figura A.10: Distribuição *a posteriori* dos parâmetros referente ao modelo III, com um ponto de mudança. A distribuição *a posteriori* corresponde a linha cheia e a distribuição *a priori* a linha tracejada. O verdadeiro valor de cada parâmetro é representado pela linha vertical.

A.4 Figuras dados simulados, modelos com dois pontos de mudança

Na Figura (A.11), mostramos a convergência das cadeias para os parâmetros do modelo II, com dois pontos de mudança, para os dados simulados.

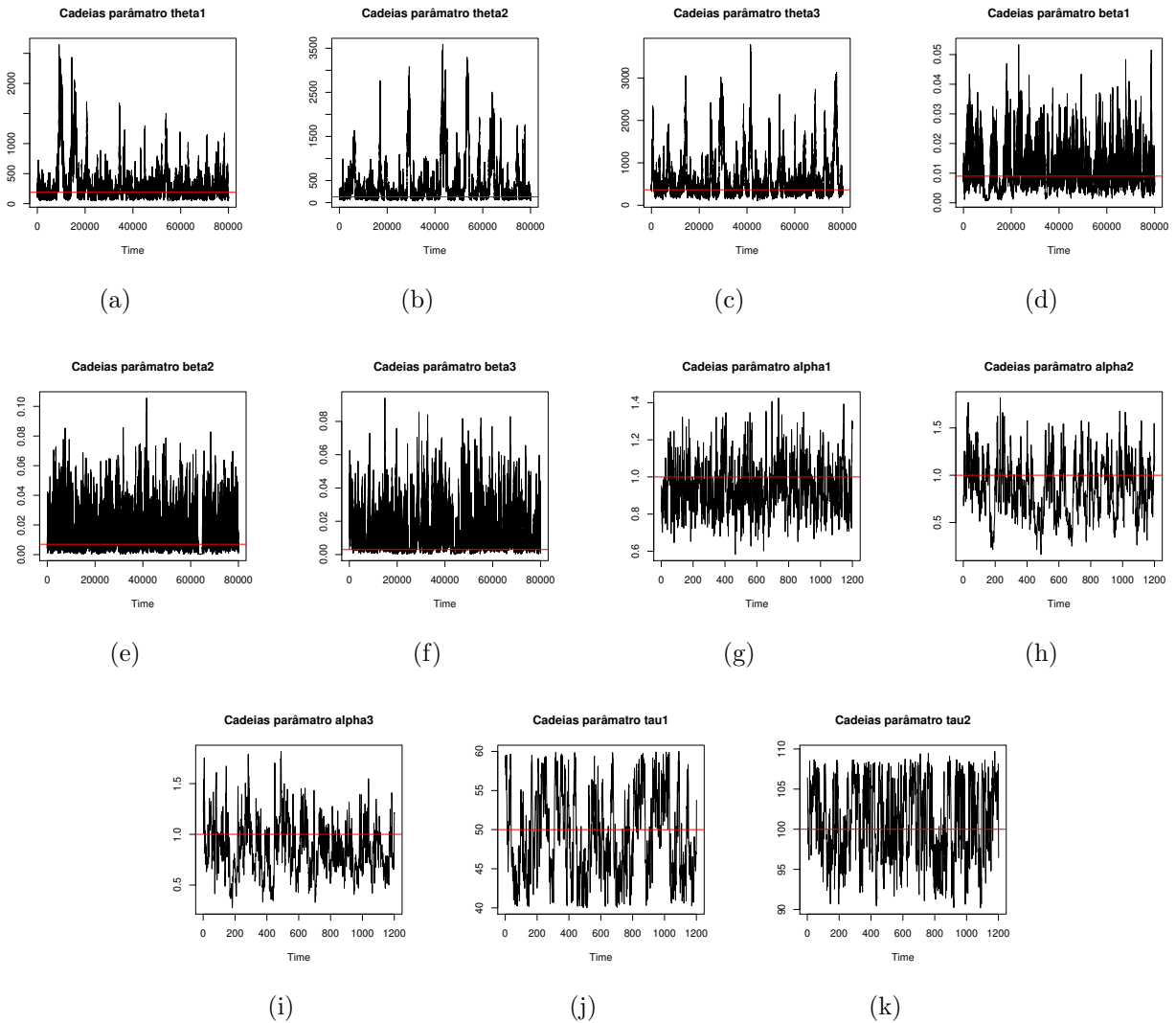


Figura A.11: Convergência das cadeias para os parâmetros do modelo II, com dois pontos de mudança, sendo que o traço na horizontal é o valor verdadeiro de cada parâmetro.

Na Figura (A.12), mostramos as distribuições *a posteriori*, *a priori* e o valor verdadeiro de cada parâmetro, para o Modelo II, com dois pontos de mudança, para os dados simulados.

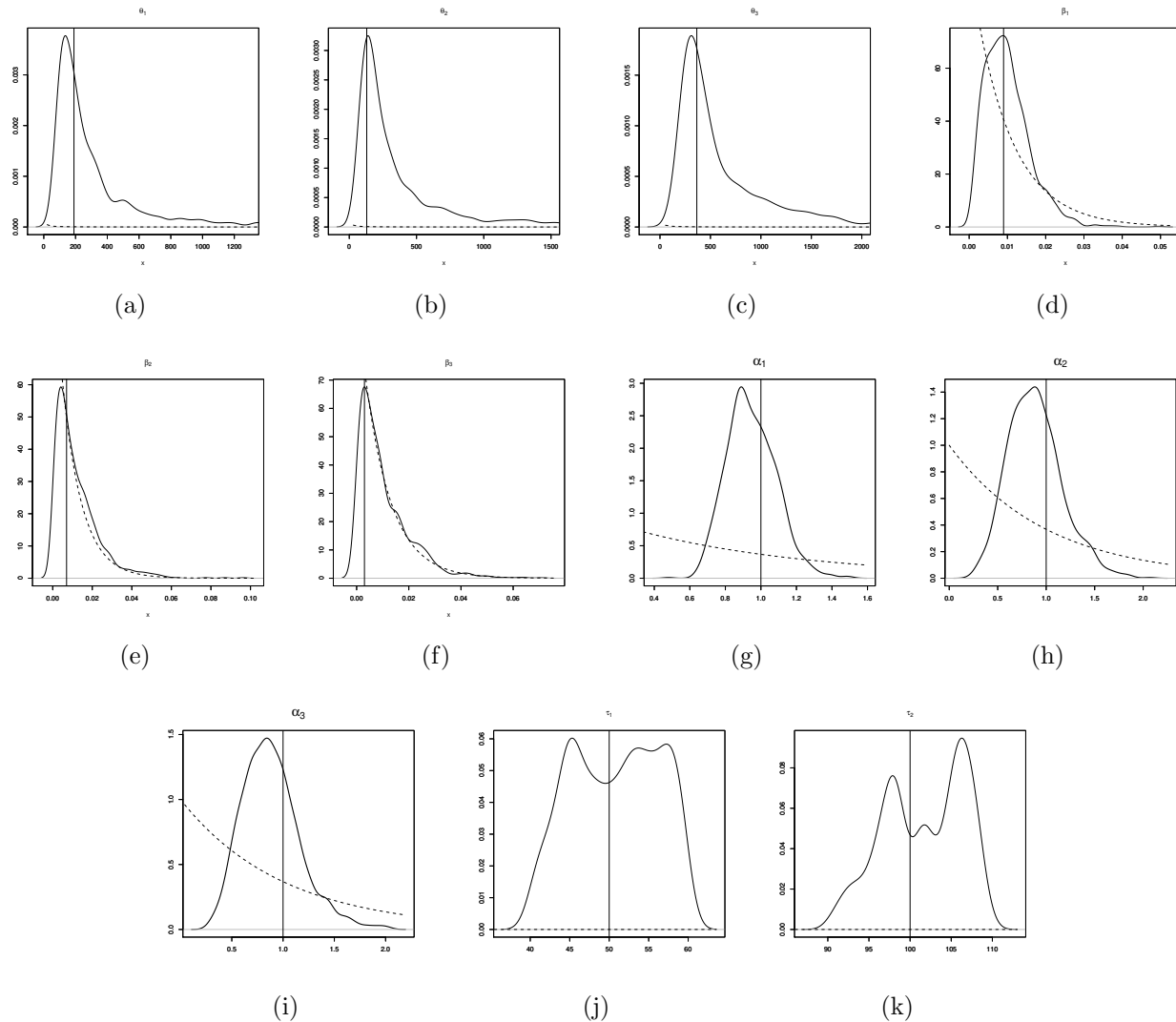


Figura A.12: Distribuição *a posteriori* dos parâmetros referente ao modelo II, com dois pontos de mudança. A distribuição *a posteriori* corresponde a linha cheia e a distribuição *a priori* a linha tracejada. O verdadeiro valor de cada parâmetro é representado pela linha vertical.

Na Figura (A.13), mostramos a convergência das cadeias para os parâmetros do Modelo III, com dois pontos de mudança, para os dados simulados.

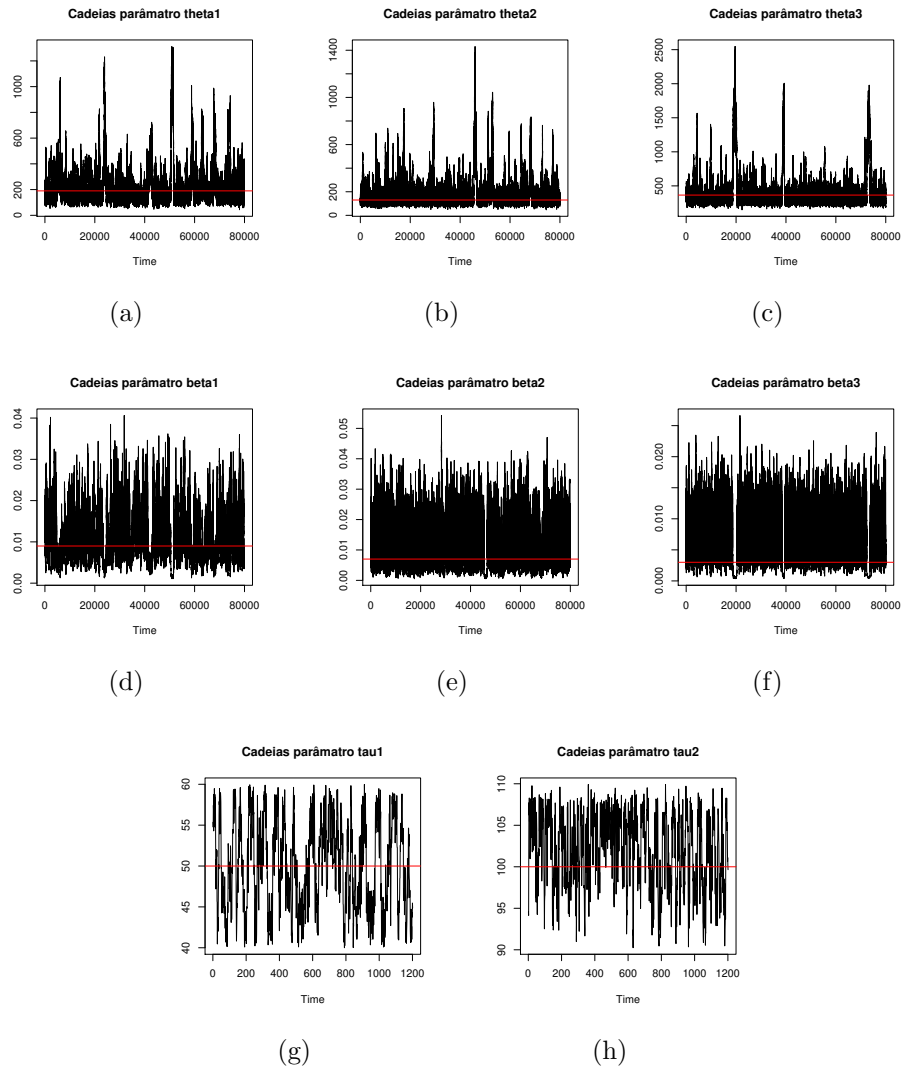


Figura A.13: Convergência das cadeias para os parâmetros do modelo III, com dois pontos de mudança, sendo que o traço na horizontal é o valor verdadeiro de cada parâmetro.

Na Figura (A.14), mostramos as distribuições *a posteriori*, *a priori* e o valor verdadeiro de cada parâmetro, para o modelo III, com dois pontos de mudança, para os dados simulados.

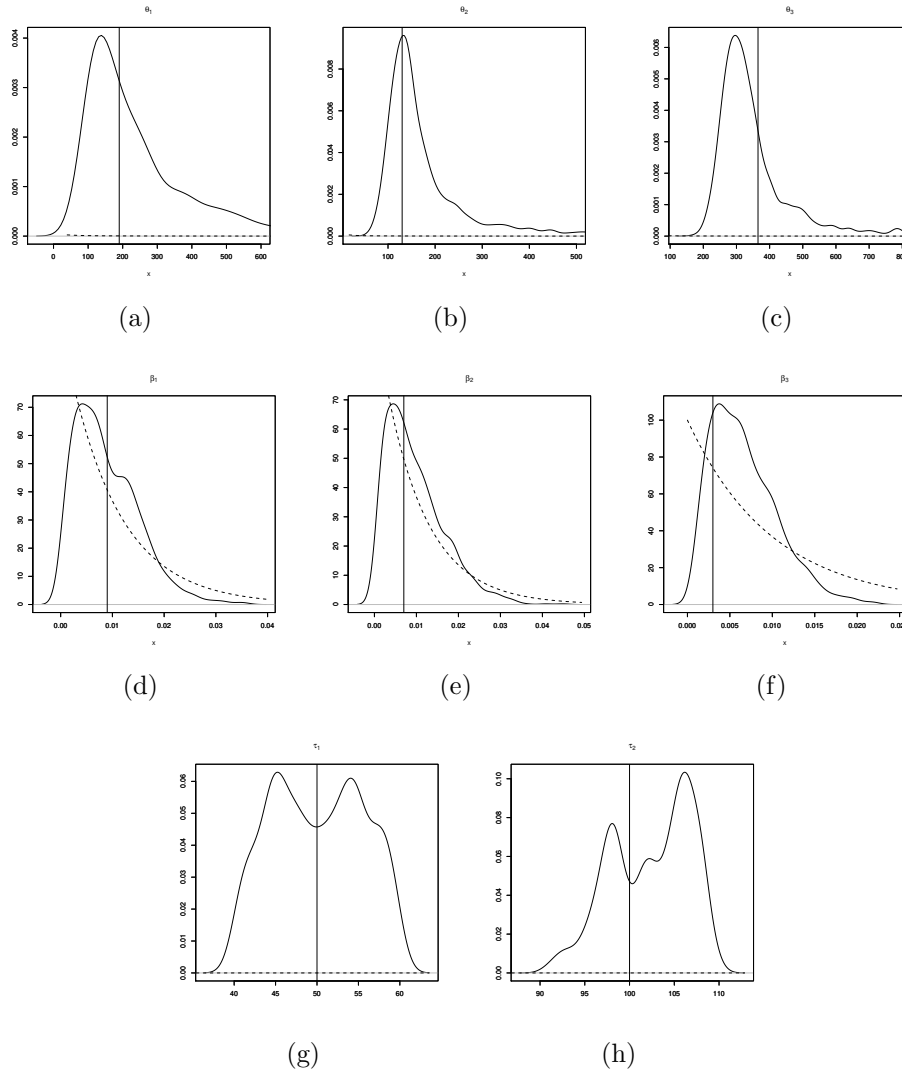


Figura A.14: Distribuição *a posteriori* dos parâmetros referente ao modelo III, com dois pontos de mudança. A distribuição *a posteriori* corresponde a linha cheia e a distribuição *a priori* a linha tracejada. O verdadeiro valor de cada parâmetro é representado pela linha vertical.

A.5 Figuras dados NE, modelos sem pontos de mudança

Na Figura (A.15), mostramos a convergência das cadeias para os parâmetros do Modelo I sem pontos de mudança, para os dados da região NE.

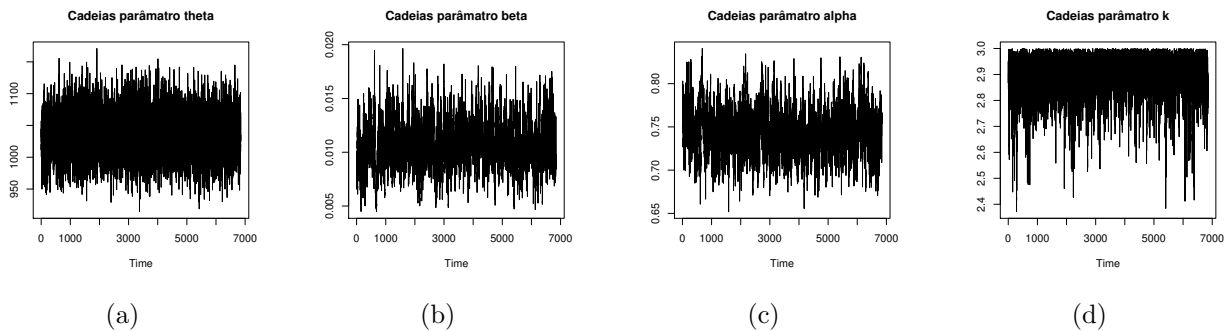


Figura A.15: Convergência dos parâmetros, modelo geral, sem pontos de mudança, dados NE.

Na Figura (A.16), mostramos as distribuições *a posteriori*, *a priori* e o valor verdadeiro de cada parâmetro, para o Modelo I sem pontos de mudança, para os dados da região NE.

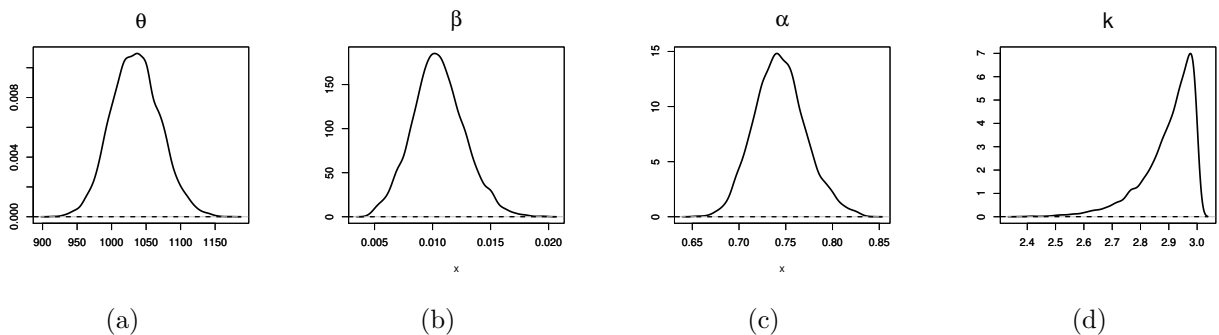


Figura A.16: Distribuição *a posteriori* dos parâmetros, modelo I, sem pontos de mudança, dados NE.

A.6 Figuras dados NE, modelos com um ponto de mudança

Na Figura (A.17), mostramos a convergência das cadeias para os parâmetros do Modelo I, com um ponto de mudança, para os dados da região NE.

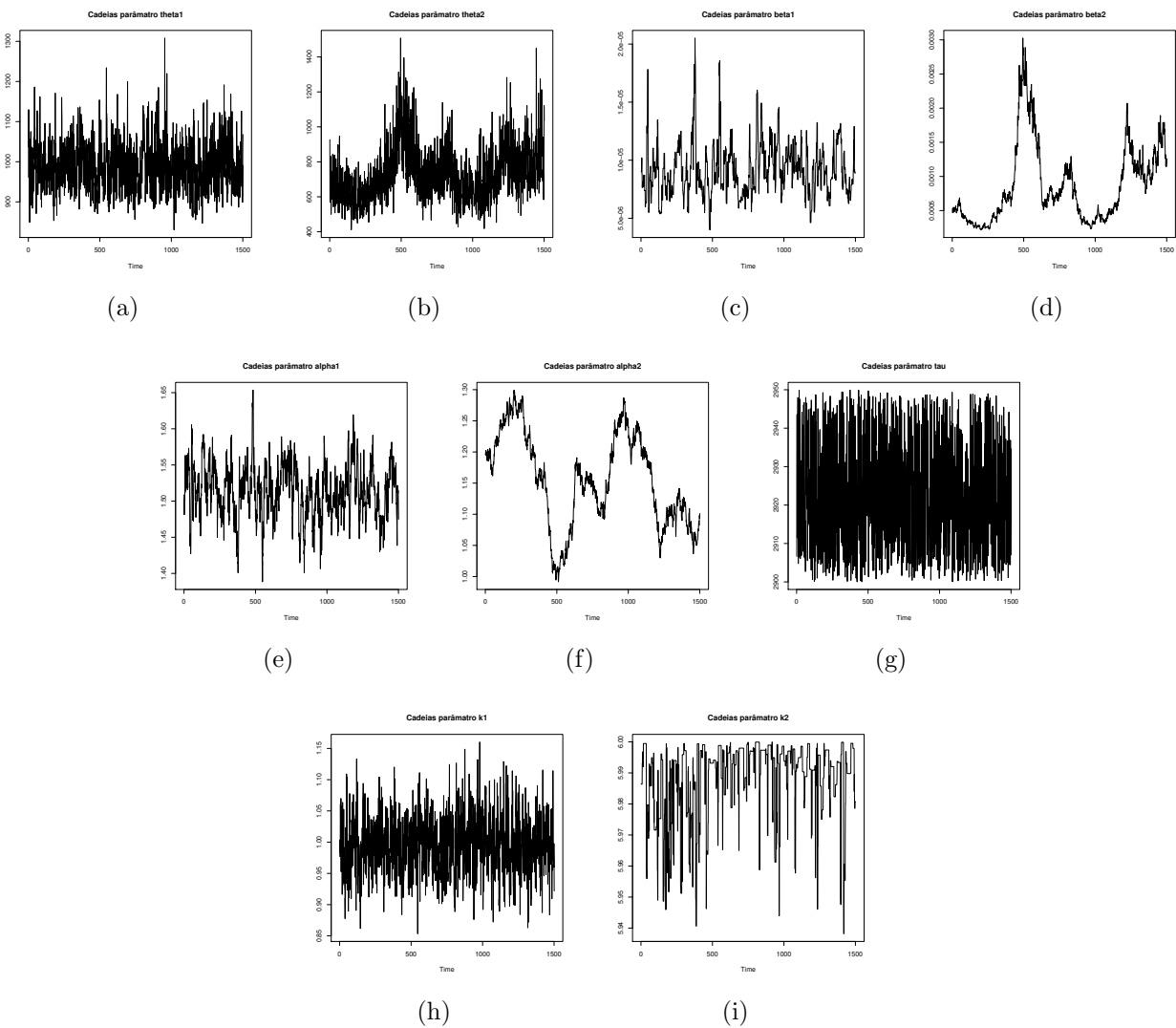


Figura A.17: Convergência dos parâmetros, para o Modelo I, com um ponto de mudança, dados NE.

Na Figura (A.18), mostramos as distribuições *a posteriori*, *a priori* de cada parâmetro, para o Modelo I, com um ponto de mudança, para os dados da região NE.

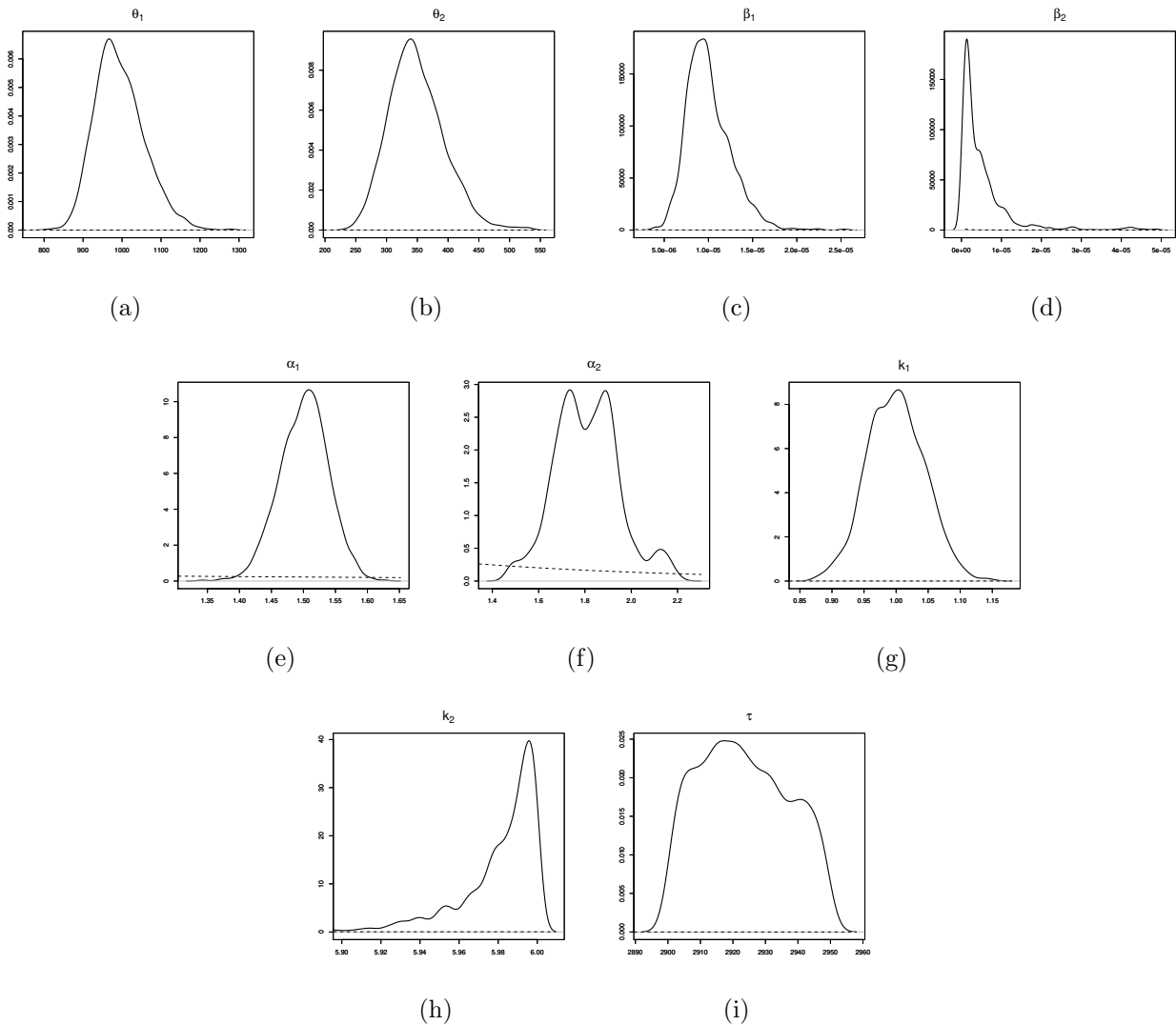


Figura A.18: Distribuição *a posteriori* dos parâmetros, Modelo I, com um ponto de mudança, para os dados da região NE. .

Na Figura (A.19), mostramos as distribuições *a posteriori*, *a priori*, para o Modelo I, com dois pontos de mudança, para os dados da região NE.

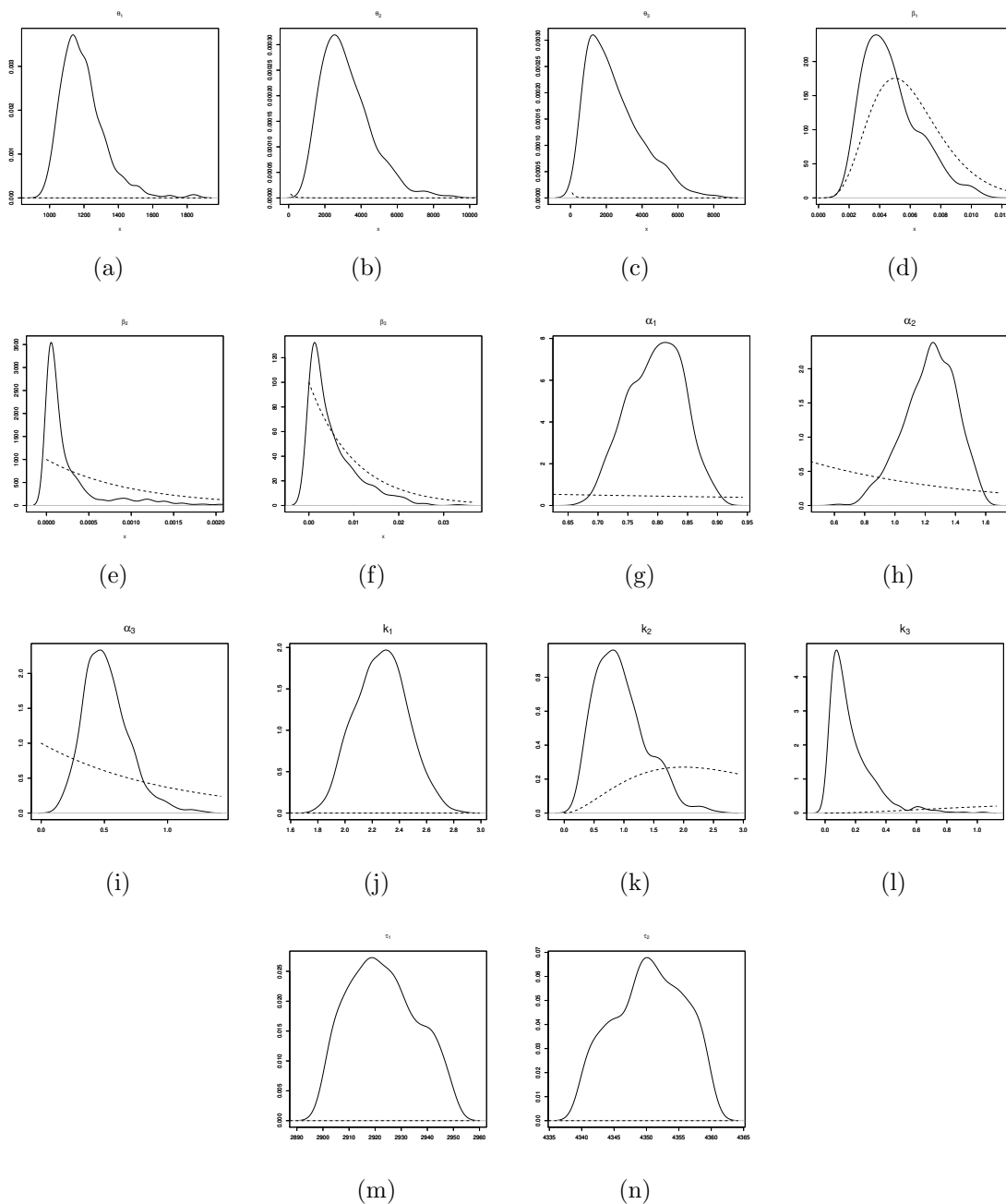


Figura A.19: Distribuição *a posteriori* dos parâmetros, referente ao Modelo I, com dois pontos de mudança, para os dados da região NE.